# This Page Is Inserted by IFW Operations and is not a part of the Official Record

# **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

# IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

### (19) World Intellectual Property Organization International Bureau





# (43) International Publication Date 4 July 2002 (04.07.2002)

# **PCT**

# (10) International Publication Number WO 02/052015 A2

- (51) International Patent Classification<sup>7</sup>: C12N 15/51, 15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85
- (21) International Application Number: PCT/CA01/01843
- (22) International Filing Date:

20 December 2001 (20.12.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/257,857

22 December 2000 (22.12.2000) US

- (71) Applicant (for all designated States except US): BOEHRINGER INGELHEIM (CANADA) LTD. [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KUKOLJ, George [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).

- (74) Agent: BERNIER, Louise, G.; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

 without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

	CLONE APGK-12	AMINO ACID SUBSTITUTIONS								
	5' HCV		EMCV			V N32→5B				3'HCV
	IRES	NeoR:	IRES	NS2	NS3	4A	NS4B	NS5A	NS5B	UTR
77 charleg	G (nt1) SEQ ID NO 1									
	A (nt1) SEQ ID NO 24				•	-	-	-	-	
86 cfu/µg	R3 rep A(nt1) SEQ ID NO 25				R(1135)K S(1560)G	K(1691)R	•	T(1993)A G(2042)C L(2155)P P(2166)L		
2000000cfu/µg	G(nt1) SEQ ID NO 7				R(1135)K S(1560)G	K(1891)R	-	T(1993)A G(2042)C L(2155)P P(2166)L		

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G-->A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

1

#### SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

#### FIELD OF THE INVENTION

The present invention relates generally to a HCV RNA molecule that self-replicates in appropriate cell lines, particularly to a self-replicating HCV RNA construct having an enhanced efficiency of establishing cell culture replication.

#### **BACKGROUND OF THE INVENTION**

disease have yet to be established.

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion and community-acquired non-A non-B hepatitis worldwide. It is estimated that over 200 million people worldwide are infected by the virus. A high percentage of carriers become chronically infected and many progress to chronic liver disease, so called chronic hepatitis C. This group is in turn at high risk for serious liver disease such as liver cirrhosis, hepatocellular carcinoma and terminal liver disease leading to death. The mechanism by which HCV establishes viral persistence and causes a high rate of chronic liver disease has not been thoroughly elucidated. It is not known how HCV interacts with and evades the host immune system. In addition, the roles of cellular and humoral immune responses in protection against HCV infection and

20

25

15

Various clinical studies have been conducted with the goal of identifying pharmaceutical compounds capable of effectively treating HCV infection in patients afflicted with chronic hepatitis C. These studies have involved the use of interferonalpha, alone and in combination with other antiviral agents such as ribavirin. Such studies have shown that a substantial number of the participants do not respond to these therapies, and of those that do respond favorably, a large proportion were found to relapse after termination of treatment. To date there are no broadly effective antiviral compounds for treatment of HCV infection.

30 HCV is an enveloped positive strand RNA virus in the *Flaviviridae* family. The single strand HCV RNA genome is of positive polarity and comprises one open reading frame (ORF) of approximately 9600 nucleotides in length, which encodes a linear polyprotein of approx. 3010 amino acids. In infected cells, this polyprotein is cleaved at multiple sites by cellular and viral proteases to produce structural and non-

35 structural (NS) proteins. The structural proteins (C, E1, E2 and E2-p7) comprise

2

polypeptides that constitute the virus particle (Hijikata et al., 1991; Grakoui et al., 1993(a)). The non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B) encode for enzymes or accessory factors that catalyze and regulate the replication of the HCV RNA genome. Processing of the structural proteins is catalyzed by host cell proteases (Hijikata et al., 1991). The generation of the mature non-structural proteins is catalyzed by two virally encoded proteases. The first is the NS2/3 zincdependent metalloprotease which auto-catalyses the release of the NS3 protein from the polyprotein. The released NS3 contains a N-terminal serine protease domain (Grakoui et al., 1993(b); Hijikata et al., 1993) and catalyzes the remaining cleavages from the polyprotein. The released NS4A protein has at least two roles. First, forming a stable complex with NS3 protein and assisting in the membrane localization of the NS3/NS4A complex (Kim et al., Arch Virol. 1999, 144: 329-343) and second, acting as a cofactor for NS3 protease activity. This membraneassociated complex, in turn catalyzes the cleavage of the remaining sites on the polyprotein, thus effecting the release of NS4B, NS5A and NS5B (Bartenschlager et al., 1993; Grakoui et al., 1993(a); Hijikata et al., 1993; Love et al., 1996; reviewed in Kwong et al., 1998). The C-terminal segment of the NS3 protein also harbors nucleoside triphosphatase and RNA helicase activity (Kim et al., 1995). The function of the protein NS4B is unknown. NS5A, a highly phosphorylated protein, seems to be responsible for the Interferon resistance of various HCV genotypes (Gale Jr. et al. 1997 Virology 230, 217; Reed et al., 1997. NS5B is an RNA-dependent RNA polymerase (RdRp) that is involved in the replication of HCV.

10

15

20

25

30

35

The open reading frame of the HCV RNA genome is flanked on its 5' end by a non-translated region (NTR) of approx. 340 nucleotides that functions as the internal ribosome entry site (IRES), and on its 3' end by a NTR of approximately 230 nucleotides. Both the 5' and 3' NTRs are important for RNA genome replication. The genomic sequence variance is not evenly distributed over the genome and the 5'NTR and parts of the 3'NTR are the most highly conserved portions. The authentic, highly conserved 3'NTR is the object of US patent 5,874,565 granted to Rice et al.

The cloned and characterized partial and complete sequences of the HCV genome have also been analyzed with regard to appropriate targets for a prospective antiviral therapy. Four viral enzyme activities provide possible targets such as (1) the NS2/3 protease; (2) the NS3/4A protease complex, (3) the NS3 Helicase and (4) the NS5B

3

RNA-dependent RNA polymerase. The NS3/4A protease complex and the NS3 helicase have already been crystallized and their three-dimensional structure determined (Kim *et al.*, 1996; Yem *et al.*, 1998; Love *et al.*, 1996; Kim *et al.*, 1998; Yao *et al.*, 1997; Cho *et al.*, 1998). The NS5B RNA dependent RNA polymerase has also been crystallized to reveal a structure reminiscent of other nucleic acid polymerases (Bressanelli *et al.* 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039; Ago *et al.* 1999, Structure 7: 1417-1426; Lesburg *et al.* 1999, Nat. Struct. Biol. 6: 937-943).

Even though important targets for the development of a therapy for chronic HCV infection have been defined with these enzymes and even though a worldwide intensive search for suitable inhibitors is ongoing with the aid of rational drug design and HTS, the development of therapy has one major deficiency, namely the lack of cell culture systems or simple animal models, which allow direct and reliable
 propagation of HCV viruses. The lack of an efficient cell culture system is still the main reason to date that an understanding of HCV replication remains elusive.

Although flavi- and pestivirus self-replicating RNAs have been described and used for the replication in different cell lines with a relatively high yield, similar experiments with HCV have not been successful to date (Khromykh *et al.*, 1997; Behrens *et al.*, 1998; Moser *et al.*, 1998). It is known from different publications that cell lines or primary cell cultures can be infected with high-titer patient serum containing HCV (Lanford *et al.* 1994; Shimizu *et al.* 1993; Mizutani *et al.* 1996; Ikda *et al.* 1998; Fourner *et al.* 1998; Ito *et al.* 1996). However, these virus-infected cell lines or cell cultures do not allow the direct detection of HCV-RNA or HCV antigens.

20

25

30

35

It is also known from the publications of Yoo et al. 1995; and of Dash et al., 1997; that hepatoma cell lines can be transfected with synthetic HCV-RNA obtained through *in vitro* transcription of the cloned HCV genome. In both publications the authors started from the basic idea that the viral HCV genome is a plus-strand RNA functioning directly as mRNA after being transfected into the cell, permitting the synthesis of viral proteins in the course of the translation process, and so new HCV particles could form HCV viruses and their RNA detected through RT-PCR. However the published results of the RT-PCR experiments indicate that the HCV replication in the described HCV transfected hepatoma cells is not particularly

4

efficient and not sufficient to measure the quality of replication, let alone measure the modulations in replication after exposure to potential antiviral drugs. Furthermore it is now known that the highly conserved 3' NTR is essential for the virus replication (Yanagi et al., 1999). This knowledge strictly contradicts the statements of Yoo et al. (supra) and Dash et al. (supra), who used for their experiments only HCV genomes with shorter 3' NTRs and not the authentic 3' end of the HCV genome.

In WO 98/39031, Rice *et al.* disclosed authentic HCV genome RNA sequences, in particular containing: a) the highly conserved 5'-terminal sequence "GCCAGCC"; b) the HCV polyprotein coding region; and c) 3'-NTR authentic sequences.

In WO 99/04008, Purcell *et al.* disclosed an HCV infectious clone that also contained only the highly conserved 5'-terminal sequence "GCCAGC".

Recently Lohman et al. 1999 (Science 285: 110-113) and Bartenschlager et al. (in CA 2,303,526, laid-open on October 3, 2000) disclosed a HCV cell culture system where the viral RNA (I377/NS2-3') self-replicates in the transfected cells with such efficiency that the quality of replication can be measured with accuracy and reproducibility. The Lohman and Bartenschlager disclosures were the first demonstration of HCV RNA replication in cell culture that was substantiated through direct measurement by Northern blots. This replicon system and sequences disclosed therein highlight once again the conserved 5' sequence "GCCAGC". A similar observation highlighting the conservation of the 5'NTR was made by Blight et al. 2000 (Science 290: 1972-1974) and WO 01/89364 published on Nov. 29, 2001.

25

30

35

5

10

In addition to the conservation of the 5' and 3' untranslated regions in cell culture replicating RNAs, three other publications by Lohman *et al.* **2001**, Krieger *et al.* **2001** and Guo *et al.* **2001** have recently disclosed distinct adaptive mutants within the HCV non-structural protein coding region. Specific nucleotide changes that alter the amino acids of the HCV non-structural proteins are shown to enhance the efficiency of establishing stable replicating HCV subgenomic replicons in culture cells.

Applicant has now found that, contrary to all previous reports, the highly conserved 5'-NTR can be mutated by adaptation to give rise to a HCV RNA sequence that, in conjunction with mutations in the HCV non-structural region, provides for a greater

5

efficiency of transduction and/or replication.

Applicant has also identified novel adaptive mutations within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture.

One advantage of the present invention is to provide an alternative to these existing systems comprising a HCV RNA molecule that self-replicates. Moreover, the present invention demonstrates that the initiating nucleotide of the plus-strand genome can be either an A as an alternative to the G already disclosed.

A further advantage of the present invention is to provide a unique HCV RNA molecule that transduces and/or replicates with higher efficiency. The Applicant demonstrates the utility of this specific RNA molecule in a cell line and its use in evaluating a specific inhibitor of HCV replication.

#### SUMMARY OF THE INVENTION

In a first embodiment, the present invention provides a 5'-non translated region of the hepatitis C virus wherein its highly conserved guanine at position 1 is substituted for adenine.

Particularly, the present invention provides a hepatitis C virus polynucleotide comprising adenine at position 1 as numbered according to the I377/NS2-3' construct (Lohmann et al. 1999, Accession # AJ242651).

Particularly, the invention provides a HCV self-replicating polynucleotide comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO. 8).

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

10

15

20

25

6

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

More particularly, the invention provides a HCV self-replicating polynucleotide encoding a polyprotein comprising a G2042C or a G2042R mutation.

Most particularly, the invention provides for HCV self-replicating polynucleotide comprising a nucleotide substitution G—>A at position 1, and said polynucleotide encodes a polyprotein further comprising a G2042C or a G2042R mutation.

Particularly, the polynucleotide of the present invention can be in the form of RNA or DNA that can be transcribed to RNA.

In a third embodiment, the invention also provides for an expression vector comprising a DNA form of the above polynucleotide, operably linked with a promoter.

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or the vector as described above.

20

10

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- incubating the host cell as described above in the absence or presence of a potential hepatitis C virus inhibitor;
- isolating the total cellular RNA from the cells;
  - analyzing the RNA so as to measure the amount of HCV RNA replicated;
  - comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.
- In a sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:
  - a) treating the above described host cell with the compound;
  - b) evaluating the treated host cell for reduced replication, wherein reduced replication indicates the ability of the compound to inhibit replication.

7

# **DETAILED DESCRIPTION OF THE DRAWINGS**

10

Figure 1 is a schematic view of the bi-cistronic replicon RNA. The sequence deviations between the I377/NS2-3' replicon from Lohman *et al.*, 1999 and the APGK12 replicon are indicated below the replicon. In place of a G nucleotide at the +1 position in the I377/NS2-3'replicon, the APGK12 contains an additional G resulting in GG at the 5' terminus (the first G being counted as position –1). In the linker region between the neo gene and the EMCV IRES sequence two areas deviate from I377/NS2-3': 14 nucleotides (CGCGCCCAGATGTT) which are not present in I377/NS2/3 are inserted at position 1184 in APGK12; 11 nucleotides (1231-1241) present in I377/NS2-3' are deleted to generate APGK-12. In the NS5B coding region, a T at position 8032 was mutated to C to eliminate a Ncol restriction site.

- 15 Figure 2 shows Northern blots of RNA-transfected Huh-7 cell lines. 12 µg of total cellular RNA or control RNA was separated on 0.5% agarose-formaldehyde gels and transferred to Hybond N+ paper, fixed and (Figure 2A) radioactively probed with HCV specific minus-strand RNA that detects the presence of plus-strand replicon RNA. Lanes 1 and 2: positive controls that contain 10<sup>9</sup> copies of in vitro transcribed APGK12 RNA. Lane 3: negative control of total cellular RNA from untransfected 20 Huh-7 cells. Lanes 4 and 5: cellular RNA from B1 and B3 cell lines that have integrated DNA copies of the neomycin phosphotransferase gene. Lane 6: total cellular RNA from a Huh-7 cell line, designated S22.3, that harbors high copy number HCV sub-genomic replicon RNA as highlighted by the arrow. Other cell lines have no detectable replicon RNA. Figure 2B is identical to Figure 2A with the 25 exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA. Lanes 1 and 2 are positive control lanes that contain 10° copies of full length HCV minus strand RNA. Lane 6, which contains 12 µg of total cellular RNA from cell line S22.3, harbors detectable minusstrand replicon RNA at the expected size of 8 - 9 kilobases. M represent the 30 migration of non-radioactive molecular size markers on the agarose gel. 28s represents the migration of 28s ribosomal RNA and accounts for the detection of this species in a samples of total cellular RNA.
- 35 Figure 3 shows indirect immunofluorescence of a HCV non-structural protein in the

S22.3 cell line. Indirect immunofluorescence was performed on cells that were cultured and fixed, permeabilized and exposed to a rabbit polyclonal antibody specific for a segment of the HCV NS4A protein. Secondary goat anti-rabbit antibody conjugated with red-fluor Alexa 594 (Molecular Probes) was used for detection. Top panels shows the results of immunofluorescence (40X objective) and the specific staining of the S22.3 cells. The bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have background level of staining.

Figure 4 shows Western-blots following SDS-PAGE separation of total proteins extracted from three cell lines: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary \ goat anti-rabbit antibody. Panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than control B1 cells and that the naïve Huh-7 cell line does not produce the NPT protein. Panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. M represents molecular weight (in kilodaltons) of pre-stained polypeptide markers.

Figure 5A and 5B identify the nucleotide and amino acid sequences respectively that differ from the APGK12 sequence in the different HCV bi-cistronic replicons. The S22.3 adapted replicon is a first generation replicon selected following the transfection of RNA transcribed from the APGK12 template. R3, R7, R16 are second generation replicons that were selected following the transfection of RNA isolated from the S22.3 first generation replicon cell line. Figure 5A: Nucleotide mutations that were characterized in each of the adapted replicons are indicated adjacent to the respective segment of the replicon (IRES, NS3, NS4A, NS5A, and NS5B). Figure 5B: Amino acid numbers are numbered according to the full length HCV poly-protein with the first amino acid in the second cistron corresponding to amino acid 810 in NS2 of I377/NS2-3' construct.

9

Figure 6 depicts the colony formation efficiency of four in vitro transcribed HCV subgenomic bi-cistronic replicon RNAs. The APGK12 serves as the reference sequence; highlighted are the initiating nucleotides of the HCV IRES in each of the constructs and the amino acid differences (from the APGK12 reference sequence) in the HCV non-structural region for the two R3-rep. Note that the in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in panels A and B) following selection with 0.25 mg/ml G418. RNA isolated from the second generation R3 cell line was reverse transcribed into DNA and cloned into the pAPGK12 vector backbone to generate the R3-rep, which was sequenced and found to encode additional changes that included the L(2155)P substitution in the NS5A segment of the HCV polyprotein (compare R3-rep sequence with the R3 sequence in tables 2 and 3). Various quantities of in vitro transcribed R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10<sup>6</sup> cfu/µg of RNA (panel C). Various quantities of R3-rep-5'G were also transfected resulting in a colony formation efficiency of 2 X 10<sup>6</sup> cfu/μg of RNA (panel D).

Figure 7 displays a typical RT-PCR amplification plot (left panel) and the graphical representation of Ct values versus known HCV RNA quantity in a standard curve (right panel). Each of the plotted curves in the left panel, graph the increment of fluorescence reporter signal (delta-Rn) versus PCR cycle number for a predetermined quantity of HCV replicon RNA. The Ct value is obtained by determining the point at which the fluorescence exceeds an arbitrary value (horizontal line). The right panel demonstrates the linear relationship between starting RNA copy number of the predetermined standards (large black dots) and the Ct value. Smaller dots are the Ct values of RNA samples (containing unknown quantity of HCV replicon RNA) from S22.3 cells treated with various concentrations of a specific inhibitor of HCV replication.

30

35

10

15

20

25

**Figure 8** shows the effect of increasing concentration of inhibitor A on HCV RNA replicon levels in Huh7 cells. S22.3 cells were grown in the presence of increasing concentrations of inhibitor A starting at 0.5nM and ranging to 1024nM. The inhibitor dose-response curve is the result of 11 concentrations from serial two-fold dilutions (1:1). One control well, without any inhibitor, was also included during the course of

10

the experiment. The cells were incubated for 4 days in a 5% CO₂ incubator at 37 °C. Total cellular RNA was extracted, quantified by optical density. HCV replicon RNA was evaluated by real time RT-PCR and plotted as genome equivalents/μg total RNA as a function of inhibitor concentration

5

10

15

35

#### **Definitions**

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell culture, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook *et al.* (1989) and Ausubel *et al.* (1994).

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission (1972).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

- The term "DNA segment or molecule or sequence", is used herein, to refer to molecules comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C). These segments, molecules or sequences can be found in nature or synthetically derived. When read in accordance with the genetic code, these sequences can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.
  - As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. The polypeptide can be encoded by a full-length sequence or any portion of the coding sequence, so long as the functional activity of the protein is retained.
- A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific structural function that constitute the viral particles. "Structural proteins" defines the HCV proteins incorporated into the virus particles namely, core "C", E1, E2, and E2-p7.
  - "Non-structural proteins", defines the HCV proteins that are not comprised in viral particles namely, NS2, NS3, NS4A, NS5A and NS5B.

11

"Restriction endonuclease or restriction enzyme" is an enzyme that has the capacity to recognize a specific base sequence (usually 4, 5 or 6 base pairs in length) in a DNA molecule, and to cleave the DNA molecule at every place where this sequence appears. An example of such an enzyme is *EcoRI*, which recognizes the base sequence G↓AATTC and cleaves a DNA molecule at this recognition site.

- "Restriction fragments" are DNA molecules produced by the digestion of DNA with a restriction endonuclease. Any given genome or DNA segment can be digested by a particular restriction endonuclease into at least two discrete molecules of restriction fragments.
- "Agarose gel electrophoresis" is an analytical method for fractionating polynucleotide molecules based on their size. The method is based on the fact that nucleic acid molecules migrate through a gel as through a sieve, whereby the smallest molecule has the greatest mobility and travels the farthest through the gel. The sieving characteristics of the gel retards the largest molecules such that, these have the least mobility. The fractionated polynucleotides can be visualized by staining the gel using methods well known in the art, nucleic acid hybridization or by tagging the fractionated molecules with a detectable label. All these methods are well known in the art, specific methods can be found in Ausubel *et al.* (*supra*).
- "Oligonucleotide or oligomer" is a molecule comprised of two or more

  deoxyribonucleotides or ribonucleotides, preferably more than three. The exact size
  of the molecule will depend on many factors, which in turn depend on the ultimate
  function or use of the oligonucleotide. An oligonucleotide can be derived
  synthetically, by cloning or by amplification.
- "Sequence amplification" is a method for generating large amounts of a target
  sequence. In general, one or more amplification primers are annealed to a nucleic
  acid sequence. Using appropriate enzymes, sequences found adjacent to, or in
  between the primers are amplified. An amplification method used herein is the
  polymerase chain reaction (PCR) and can be used in conjunction with the reversetranscriptase (RT) to produce amplified DNA copies of specific RNA sequences.
- "Amplification primer" refers to an oligonucleotide, capable of annealing to a RNA or
   DNA region adjacent to a target sequence and serving as the initiation primer for
   DNA synthesis under suitable conditions well known in the art. The synthesized
   primer extension product is complementary to the target sequence.
  - The term "domain" or "region" refers to a specific amino acid sequence that defines either a specific function or structure within a protein. As an example herein, is the

35

12

NS3 protease domain comprised within the HCV non-structural polyprotein. The terms "plasmid" "vector" or "DNA construct" are commonly known in the art and refer to any genetic element, including, but not limited to, plasmid DNA, phage DNA, viral DNA and the like which can incorporate the oligonucleotide sequences, or sequences of the present invention and serve as DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The terminology "expression vector" defines a vector as described above but designed to enable the expression of an inserted sequence following transformation or transfection into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. Such expression control sequences will vary depending on whether the vector is designed to express the operably linked gene *in vitro* or *in vivo* in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

10

15

20

25

30

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) or RNA, when such nucleic acid has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting/transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, an example of a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome and is inherited by daughter cells through chromosome replication. A host cell or indicator cell can be transfected with RNA. A cell can be stably transfected with RNA if the RNA replicates and copies of the RNA segregate to daughter cells upon cell division. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA or RNA. Transfection methods are well known in the art (Sambrook et al., 1989; Ausubel et al., 1994). If the RNA encodes for a genetic marker that imparts an observable phenotype, such as antibiotic resistance, then the stable transfection of replicating RNA can be monitored by the acquisition of such phenotype by the host cell.

As used herein the term "transduction" refers to the transfer of a genetic marker to host cells by the stable transfection of a replicating RNA.

13

The nucleotide sequences and polypeptides useful to practice the invention include without being limited thereto, mutants, homologs, subtypes, quasi-species, alleles, and the like. It is understood that generally, the sequences of the present invention encode a polyprotein. It will be clear to a person skilled in the art that the polyprotein of the present invention and any variant, derivative or fragment thereof, is auto-processed to an active protease.

As used herein, the designation "variant" denotes in the context of this invention a sequence whether a nucleic acid or amino acid, a molecule that retains a biological activity (either functional or structural) that is substantially similar to that of the original sequence. This variant may be from the same or different species and may be a natural variant or be prepared synthetically. Such variants include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided the biological activity of the protein is conserved. The same applies to variants of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained.

10

15

20

30

35

The term "derivative" is intended to include any of the above described variants when comprising additional chemical moiety not normally a part of these molecules. These chemical moieties can have varying purposes including, improving a molecule's solubility, absorption, biological half life, decreasing toxicity and eliminating or decreasing undesirable side effects. Furthermore, these moieties can be used for the purpose of labeling, binding, or they may be comprised in fusion product(s). Different moieties capable of mediating the above described effects can be found in *Remington's The Science and Practice of Pharmacy* (1995).

25 Methodologies for coupling such moieties to a molecule are well known in the art.

The term "fragment" refers to any segment of an identified DNA, RNA or amino acid sequence and/or any segment of any of the variants or derivatives described herein above that substantially retains its biological activity (functional or structural) as required by the present invention.

The terms "variant", "derivative", and "fragment" of the present invention refer herein to proteins or nucleic acid molecules which can be isolated/purified, synthesized chemically or produced through recombinant DNA technology. All these methods are well known in the art. As exemplified herein below, the nucleotide sequences and polypeptides used in the present invention can be modified, for example by *in vitro* mutagenesis.

:

14

As used herein, the term "HCV polyprotein coding region" means the portion of a hepatitis C virus that codes for the polyprotein open reading frame (ORF). This ORF may encode proteins that are the same or different than wild-type HCV proteins. The ORF may also encode only some of the functional protein encoded by wild-type polyprotein coding region. The protein encoded therein may also be from different isolates of HCV, and non-HCV protein may also be encoded therein.

As used herein, the abbreviation "NTR" used in the context of a polynucleotide molecule means a non-translated region. The term "UTR" means untranslated region. Both are used interchangeably.

#### Preferred embodiments

10

20

25

35

Particularly, the invention provides a HCV self-replicating polynucleotide molecule comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO.8).

According to the first embodiment of this invention, there is particularly provided a HCV polynucleotide construct comprising:

- a 5'-non translated region (NTR) comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein coding region; and
- a 3'-NTR region.

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

Alternatively, the first embodiment of the present invention is directed to HCV self-replicating polynucleotide molecule comprising a G2042C/R mutation.

15

According to the second embodiment, the present invention particularly provides a HCV polynucleotide construct comprising:

- a 5'-NTR region comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein region coding for a HCV polyprotein comprising a
   G(2042)C or a G(2042)R mutation; and
- a 3'-NTR region.

5

20

25

30

- Preferably, the polynucleotide construct of the present invention is a DNA or RNA molecule. More preferably, the construct is a RNA molecule. Most preferably, the construct is a DNA molecule.
- More particularly, the first embodiment of this invention is directed to a RNA

  molecule encoded by the DNA molecule selected from the group consisting of: SEQ

  ID NO. 2, 4, 5, 6, 7, 24 and 25.
  - Most particularly, the invention provides a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
  - In a third embodiment, the invention also is directed to an expression vector comprising DNA forms of the above polynucleotide, operably linked with a promoter.
  - Preferably, the promoter is selected from the group consisting of: T3, T7 and SP6.
  - According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or vector as described above. Particularly, the host cell is a eukaryotic cell line. More particularly, the eukaryotic cell line is a hepatic cell line. Most particularly, the hepatic cell line is Huh-7.
  - In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:
    - a) incubating the host cell as described above under conditions suitable for RNA replication;
- b) isolating the total cellular RNA from the cells; and

16

c) analyzing the RNA so as to measure the amount of HCV RNA replicated.

Preferably, the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

Alternatively in this fifth embodiment, the construct comprises a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.

10

15

According to a preferred aspect of the sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:

- a) carrying step a) as described in the above assay, in the presence or absence of the compound;
- b) isolating the total cellular RNA from the cells; and
- c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.
- 20 wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

Preferably, the cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

25

# **EXAMPLES**

# EXAMPLE 1

Replicon Constructs (APGK-12: Figure 1)

pET9a-EMCV was obtained by ligating an oligonucleotide linker
5' gaattccagatggcgcgcccagatgttaaccagatccatggcacactctagagtactgtcgac 3' (SEQ ID NO.9) to pET-9a (Novagen) that was cut with EcoRI and Sall to form the vector pET-9a-mod. This linker contains the following restriction sites: EcoRI, AscI, HpaI, NcoI, XbaI, ScaI, Sall. The EMCV IRES was amplified by PCR from the vector pTM1 with primers

17

5' cggaatcgttaacagaccacaacggtttccctc 3' (SEQ ID NO.10) and 5' ggcgtacccatggtattatcgtgtttttca 3' (SEQ ID NO.11) and ligated into pET-9a-mod via EcoRI and NcoI to form pET-9a-EMCV.

- 5 The sequence of HCV NS2 to NS5B followed by the 3'UTR of HCV was obtained from the replicon construct I377/NS2-3' (Lohman et al., 1999; accession number: AJ242651) and synthesized by Operon Technologies Inc. with a T to C change at the Ncol site in NS5B at nucleotide 8032. This sequence was released from an GenOp® vector (Operon Technologies) with Ncol and Scal and transferred into pET-9a-EMCV-NS2-5B-3'UTR.
  - pET-9a-HCV-neo was obtained by amplification of the HCV IRES from a HCV cDNA isolated from patient serum with primers
  - 5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.12) containing a T7 promoter and primer
    - 5' ggcgcgccctttggtttttctttgaggtttaggattcgtgctcat 3' (**SEQ ID NO.13**) and amplification of the neomycin phosphotransferase gene from the vector pcDNA 3.1 (Invitrogen) with primers
    - 5' aaagggcgcatgattgaacaagatggattgcacgca 3' (SEQ ID NO.14) and 5'
- 20 gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15). These two PCR fragments were mixed and amplified with primers
  - 5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.16) and 5' gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15); cut with Eco RI and Hpal and transferred into pET-9a-mod to form pet-9a-HCV-neo. The EMCV-
- NS2-5B-3'UTR was released from pET-9a-EMCV-NS2-5B-3'UTR with Hpal and Scal and transferred into pet-9a-HCV-neo that was cut with Hpal to form pET-9a-APGK12. This insert was sequenced with specific successive primers using a ABI Prism® BigDye™ Terminator Cycle sequencing kit and analyzed on ABI Prism® 377 DNA Sequencer and is shown in SEQ ID NO 1.

RNA in vitro transcription

30

35

pET-9a-APGK12 DNA was cut with Scal for expression of the full-length replicon or with Bglll for expression of a truncated negative control RNA. DNA was analyzed on a 1% agarose gel and purified by Phenol/Chloroform extraction. RNA was produced using a T7 Ribomax® kit (Promega) followed by extraction with phenol/chloroform

18

and precipitation with 7.5 M LiCl<sub>2</sub>. RNA was treated with DNAse I for 15 min to remove the DNA template and further purified with an RNeasy® column (Qiagen). RNA integrity was verified on a denaturing formaldehyde 1% agarose gel.

#### 5 EXAMPLE 2

10

15

20

25

30

Primary transfection of Huh7 cells and selection of replicon cell lines Human hepatoma Huh7 cells (Health Science Research Resources Bank, Osaka, Japan) were grown in 10% FBS/DMEM. Cells were grown to 70% confluency, trypsinized, washed with phosphate buffered saline (PBS) and adjusted to 1x10<sup>7</sup> cells/ml of PBS. 800 μl of cells were transferred into 0.4cm cuvettes and mixed with 15 μg of replicon RNA. Cells were electroporated using 960μF, 300 volts for ~18 msec and evenly distributed into two 15 cm tissue culture plates and incubated in a tissue culture incubator for 24 hours. The selection of first and second generation replicon cell lines was with 10% FBS/DMEM medium supplemented with 1mg/ml of G418. Cells were selected for 3-5 weeks until colonies were observed that were isolated and expanded.

Following the G418 selection and propagation of Huh-7 cells transfected with APGK12 (SEQ ID NO. 1) RNA, cells that formed a distinct colony were treated with trypsin and serially passed into larger culture flasks to establish cell lines. Approximately 10 X 10<sup>6</sup> cells were harvested from each cell line. The cells were lysed and the total cellular RNA extracted and purified as outlined in Qiagen RNAeasy® preparatory procedures. Figure 2 shows the analysis of 12 µg of total cellular RNA from various cell lines as analyzed on a Northern blot of a denaturing agarose-formaldehyde gel.

Figure 2A is a Northern blot (radioactively probed with HCV specific minus-strand RNA) that detects the presence of plus-strand replicon RNA. Lanes 1 and 2 are positive controls that contain 10<sup>9</sup> copies of in vitro transcribed APGK12 RNA. Lane 2 contains the *in vitro* transcribed RNA mixed with 12 µg of total cellular from naïve Huh-7 cells. Lane 3 is a negative control of total cellular RNA from untreated Huh-7 cells. Lanes 4 and 5 contain cellular RNA from the B1 and B3 G418 resistant cell lines that have DNA integrated copies of the neomycin phosphotransferase gene. Lane 6 contains total cellular RNA from a Huh-7 cell line, designated S22.3, that

harbors high copy number of HCV sub-genomic replicon RNA as detected by the positive signal in the 8 kilo-base range. Other cell lines have no detectable replicon RNA. Figure 2B is a Northern blot of a duplicate of the gel presented in 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA (lanes 1 and 2 are positive control lanes that contain  $10^9$  copies of full length genomic HCV minus strand RNA); only lane 6, which contains 12  $\mu$ g of total cellular RNA from cell line S22.3, harbors detectable minus-strand replicon RNA at the expected size of 8 – 9 kilobases. An quantitative estimation of RNA copy number, based on phosphorimager scanning of the Northern blots, is approximately 6 X10 $^7$  copies of plus-strand/ $\mu$ g of total RNA, and 6 x  $10^6$  copies of minus strand/ $\mu$ g of total RNA. The presence of the plus-strand and minus-strand intermediate confirms that the HCV sub-genomic RNA is actively replicating in the S22.3 cell line.

#### 15 EXAMPLE 3

10

20

25

30

35

# S22.3 cell line constitutively expresses HCV non-structural proteins.

HCV non-structural protein expression was examined in the S22.3 cell line. Figure 3 displays the result of indirect immunofluorescence that detects the HCV NS4A protein in the S22.3 cell line and not in the replicon negative B1 cell line (a G418 resistant Huh-7 cell line). Indirect immunofluorescence was performed on cells that were cultured and fixed (with 4% paraformaldehyde) onto Lab-tek chamber slides. Cells were permeabilized with 0.2% Triton X-100 for 10 minutes followed by a 1 hour treatment with 5% milk powder dissolved in phosphate-buffered saline (PBS). A rabbit serum containing polyclonal antibody raised against a peptide spanning the HCV NS4A region was the primary antibody used in detection. Following a 2 hour incubation with the primary antibody, cells were washed with PBS and a secondary goat anti-rabbit antibody conjugated with red-fluor Alexa® 594 (Molecular Probes) was added to cells for 3 hours. Unbound secondary antibody was removed with PBS washes and cells were sealed with a cover slip. Figure 3 (top panels) shows the results of immunofluorescence as detected by a microscope with specific fluorescent filtering; the bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have

20

background level of staining. A small proportion of S22.3 cells express high levels of intensely stained HCV NS4A.

Expression of the proteins encoded by the bi-cistronic replicon RNA was also examined on Western-blots following SDS-PAGE separation of total proteins extracted from: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Figure 4 panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary HRP-conjugated goat anti-rabbit antibody. Figure 4 panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than B1 cells (which contain an integrated DNA copy of the *npt* gene) and that the naïve Huh-7 cell line does not produce the NPT protein. Figure 4 panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. The western blots demonstrate that the S22.3 cell line, which harbors actively replicating HCV sub-genomic replicon RNA, maintains replication of the RNA through the high level expression of the HCV non-structural proteins.

#### 20 EXAMPLE 4

10

15

25

30

# Sequence determination of adapted replicons

Total RNA was extracted from replicon containing Huh7 cells using a RNeasy Kit (Qiagen). Replicon RNA was reverse transcribed and amplified by PCR using a OneStep RT-PCR kit (Qiagen) and HCV specific primers (as selected from the full-length sequence disclosed in WO 00/66623). Ten distinct RT-PCR products, that covered the entire bi-cistronic replicon in a staggered fashion, were amplified using oligonucleotide primers. The PCR fragments were sequenced directly with ABI Prism® BigDye™ Terminator Cycle PCR Sequencing and analyzed on ABI Prism® 377 DNA Sequencer. To analyze the sequence of the HCV replicon 3' and 5' ends a RNA ligation/RT-PCR procedure described in Kolykhalov *et al.* 1996 was followed. The nucleotide sequence of S22.3 is presented as SEQ ID NO. 2.

21

# EXAMPLE 5.

10

15

20

25

30

# Serial Passage of HCV Replicon RNA

The total cellular RNA from the S22.3 cell line was prepared as described above. HCV Replicon RNA copy number was determined by Tagman® RT-PCR analysis and 20 μg of total S22.3 cellular RNA (containing 1 X 109 copies of HCV RNA) was transfected by electroporation into 8 X 10<sup>6</sup> naïve Huh-7 cells. Transfected cells were subsequently cultured in 10 cm tissue culture plates containing DMEM supplemented with 10% fetal calf serum (10% FCS). Media was changed to DMEM (10% FCS) supplemented with 1 mg/ml G418 24 hours after transfection and then changed every three days. Twenty-three visible colonies formed three to four weeks post-transfection and G418 selection. G418 resistant colonies were expanded into second generation cell lines that represent the first cell lines harboring serially passaged HCV Replicon RNA. Three of these cell lines: R3, R7, and R16 were the subject of further analyses. First, the efficiency of transduction by each of the adapted replicons was determined by electroporation of the total cellular RNA (extracted from the R3, R7 and R16) into naïve Huh-7 cells; following electroporation, the transduction efficiency was determined as described above, by counting the visible G418 resistant colonies that arose following 3 to 5 weeks of G418 selection (Table 1). Second, the sequence of the serially passed adapted replicons was determined from the total cellular RNA that was extracted from each of the R3, R7 and R16 replicon cell lines as described in example 4 (SEQ ID NO. 4, 5, 6). Using the pAPGK12 as a reference sequence (SEQ ID NO. 1), the nucleotide changes that were selected in HCV segment of the adapted replicons are presented in Figure 5A. Some of these nucleotide changes are silent and do not change the encoded amino acid whereas others result in an amino acid substitution. Figure 5B summarizes the amino acid changes encoded by the adapted replicons with the amino acid sequence of pAPGK12 as the reference. It is important to note that the reference sequence APGK-12 (SEQ ID NO.1) contains an extra G at the 5'-terminal (5'-GG) that is not maintained in the replicating RNA of the established cell lines. Also noteworthy is that, in addition to G->A at nucleotide 1, there is also an adapted mutation G->C/R at amino acid 2042 (shown as amino acid 1233 in the sequence listing since a.a. 810 of NS2 is numbered as a.a. 1 in SEQ ID) that can be found in all clones analyzed.

TABLE 1

22

Transfection of Huh-7 cells

	RNA	Copies of Replicon	# Colonies	SEQ ID
5				
	5 ng APKG12 replicon in 20μg total Huh-7 RNA	1.2 x 10 <sup>9</sup>	0	
10	15 μg APKG12 replicon RNA	3 x 10 <sup>12</sup>	1 (\$22.3)	1
	20μg total: S22.3 cellular RNA	3 x 10 <sup>9</sup>	23 (3 clones analyzed)	2
15	R3 cellular RNA	1 x 10 <sup>9</sup>	200	4
	R7 cellular RNA	1 x 10 <sup>9</sup>	20	5
	R16 cellular RNA	3 x 10 <sup>8</sup>	100	6
	cloned R3rep RNA	2.3 x 10 <sup>8</sup>	2000	7

#### 20 EXAMPLE 6

25

30

# Construction of APGK12 with 5' G-> A substitution (APGK12-5'A, SEQ ID NO.24)

The pAPGK12 DNA was modified to change the first nucleotide in the sequence to replace the 5'GG with a 5'A. The change in the pAPGK12 was introduced by replacing an *EcoRI/AgeI* portion of the sequence with a PCR-generated *EcoRI/AgeI* fragment that includes the mutation. The oligonucleotides used for the amplification were (SEQ ID. NO. 20): 5'-GTG GAC GAA TTC TAA TAC GAC TCA CTA TAA CCA GCC CCC GAT TGG-3' and (SEQ ID. NO. 21): 5'-GGA ACG CCC GTC GTG GCC AGC CAC GAT-3' and generated a 195 bp DNA fragment that was then digested with *EcoRI* and *AgeI*. The resulting 178 bp restriction fragment was used to replace the *EcoRI* / *AgeI* fragment in pAPGK12 to generate the pAPGK12-5'A plasmid.

#### EXAMPLE 7

cDNA CLONING OF THE R3-REPLICON (R3REP).

The cDNA clone of the R3 replicon was produced by RT-PCR of RNA extracted from the R3 cell line. The following two oligonucleotides were used: (SEQ ID. NO. 22): 5'-GTC GTC TCT GAC ATG GAG AC-3' and (SEQ ID. NO. 23): 5'-GAG TTG

CTC AGT GGA TTG ATG GGC AGC-3'. The ~4400nt PCR fragment, starting within the NS2 coding region and extending to the 5'-end of the NS5B coding region, was cloned into the plasmid pCR3.1 by TA cloning (Invitrogen). The SacII / XhoI portion of this R3 sequence was then used to replace the SacII / XhoI fragment present in the pAPGK12 and the pAPGK12-5'A described above. Consequently, two R3 cDNA sequences were generated: (I) R3-Rep-5'G with an initiating 5'G (SEQ ID NO.7), and R3-Rep-5'A (SEQ ID NO.25) with an initiating 5'A. Sequencing of the R3 rep cDNA identified unique nucleotide changes that differ from the original pAPGK12 sequence (see Figure 5A); some of these changes are silent and do not change the encoded amino acid, whereas others do result in an amino acid change (see Figure 5B). The differences between R3 and the R3-rep reflect the isolation of a unique R3-rep cDNA clone encoding nucleotide changes that were not observed from the sequencing of the total RNA extracted from the R3 cell line.

#### 15 EXAMPLE 8

10

20

25

30

35

### Efficiency of colony formation with modified constructs

RNA from pAPGK12, pAPGK12-5'A, pR3-Rep and pR3-Rep-5'A was generated by in vitro transcription using the T7 Ribomax® kit (Promega) as described in example 1 above. The reactions containing the pAPGK12-5'A and pR3-Rep-5'A templates were scaled-up 10-fold due to the limitation of commercial RNA polymerase in initiating transcripts with 5'-A. The full length RNAs and control truncated RNA for each clone were introduced into 8 x 10<sup>6</sup> naïve Huh-7 cells by electroporation as described in example 2. Replicon RNA was supplemented with total cellular Huh-7 carrier RNA to achieve a final 15-20µg quantity. The cells were then cultured in DMEM medium supplemented with 10% fetal calf serum and 0.25 mg/ml G418 in two 150 mm plates. The lower concentration of G418 was sufficient to isolate and select replicon containing cell lines as none of the transfectants with the control truncated RNA produced any resistant colonies. In contrast, in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in Figure 6 panels A and B) following selection with G418. Various quantities (ranging from 0.1 ng to 1 µg) of the R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10<sup>6</sup> cfu/μg of RNA (Figure 6 panel C depicts transfection with 1 μg of RNA). Various quantities (ranging from 0.1 ng to 1 µg) of R3-rep [5'G] were similarly transfected resulting in a colony formation efficiency of 2 X 106 cfu/µg of RNA (Figure 6 panel D

24

depicts colony formation with 1μg of RNA). Note that, shown for the first time, HCV subgenomic replicons replicate as efficiently with a 5' A nucleotide in place of the 5'G. APGK12 with a 5'A or 5'G RNA have similar transduction efficiencies. Similarly, R3-Rep RNAs with either the 5'A or 5'G both display the markedly increased transduction efficiency. Notably, the adaptive mutants within the HCV non-structural segment encoded by the R3-Rep provides for a substantial increase in transduction efficiency as depicted by the dramatic increase in colony forming units per μg of transfected RNA.

# 10 EXAMPLE 9

15

20

25

# Quantification of HCV Replicon RNA Levels in Cell lines

S22.3 cells, or cell lines harboring other adapted replicons, were seeded in DMEM supplemented with 10% FBS, PenStrep and 1μg/mL Geneticin. At the end of the incubation period the replicon copy number is evaluated by real-time RT-PCR with the ABI Prism 7700 Sequence Detection System. The TAQMAN® EZ RT-PCR kit provides a system for the detection and analysis of HCV RNA (as first demonstrated by Martell *et al.* **1999** J. Clin. Microbiol. 37: 327-332). Direct detection of the reverse transcription polymerase chain reaction (RT-PCR) product with no downstream processing is accomplished by monitoring the increase in fluorescence of a dyelabeled DNA probe (Figure 6). The nucleotide sequence of both primers (adapted from Ruster, B. Zeuzem, S. and Roth, W.K., **1995**. Analytical Biochemistry 224:597-600) and probe (adapted from Hohne, M., Roeske, H. and Schreier, E. **1998**, Poster Presentation: P297 at the Fifth International Meeting on Hepatitis C Virus and Related Viruses Molecular Virology and Pathogenesis, Venezia-Lido Italy, June 25-28, 1998) located in the 5'-region of the HCV genome are the following:

HCV Forward primer:

5' ACG CAG AAA GCG TCT AGC CAT GGC GTT AGT 3' (SEQ ID NO.17)

30

HCV Reverse primer:

5' TCC CGG GGC ACT CGC AAG CAC CCT ATC AGG 3' (SEQ ID NO.18)

**HCV Probe:** 

5' FAM-TGG TCT GCG GAA CGG GTG AGT ACA CC-TAMRA 3' (**SEQ ID NO.19**)

5 FAM: Fluorescence reporter dye.

TAMRA: Quencher dye.

Using The TAQMAN® EZ RT-PCR kit, the following reaction was set up:

Component	Volume per sample	Final	
	(μ <b>L</b> )	Concentration	
RNase-Free Water	16	-	
5X Taqman EZ Buffer	10	1X	
Manganese Acetate 25mM	6	3mM	
dATP 10mM	1.5	300µM	
dCTP 10mM	1.5	300µM	
dGTP 10mM	1.5	300µM	
dUTP 20mM	1.5	300µM	
HCV Forward Primer 10µM	1	200nM	
HCV Reverse Primer 10µM	1	200nM	
HCV Probe 5uM	2	200nM	
rTth DNA Polymerase	2	0.1U/μL	
2.5U/µL			
AmpErase UNG 1U/μL	0.5	0.01U/µL	
Total Mix	45	-	

10

15

To this reaction mix,  $5\mu$ L of total RNA extracted from S22.3 cells diluted at  $10 \text{ng}/\mu$ L was added, for a total of 50 ng of RNA per reaction. The replicon copy number was evaluated with a standard curve made from known amounts of replicon copies (supplemented with 50 ng of wild type Huh-7 RNA) and assayed in an identical reaction mix (Figure 7).

Thermal cycler parameters used for the RT-PCR reaction on the ABI Prism 7700 Sequence Detection System were optimized for HCV detection:

26

Cycle	Temperature (°C)	Time (Minutes)	Repeat	Reaction
Hold	50	2	-	Initial Step
Hold	60	30		Reverse
				Transcription
Hold	95	5		<b>UNG Deactivation</b>
Cycle	95 -	0:15	2	Melt
	60	1	2	Anneal/Extend
Cycle	90	0:15	40	Melt
	60	1	40	Anneal/Extend

Quantification is based on the threshold cycle, where the amplification plot crosses a defined fluorescence threshold. Comparison of the threshold cycles provides a highly sensitive measure of relative template concentration in different samples. Monitoring during early cycles, when PCR fidelity is at its highest, provides precise data for accurate quantification. The relative template concentration can be converted to RNA copy numbers by employing a standard curve of HCV RNA with known copy number (Figure 7).

#### 10 **EXAMPLE 10**

15

20

A specific HCV NS3 protease anti-viral compound inhibits replication of the HCV replicon in S22.3 cell lines.

In order to determine the effect of a specific HCV NS3 protease anti-viral compound on replicon levels in S22.3 cells, the cells were seeded in 24 Well Cell Culture Cluster at 5 X 10<sup>4</sup> cells per well in 500μL of DMEM complemented with 10% FBS, PenStrep and 1μg/mL Geneticin. Cells were incubated until compound addition in a 5% CO<sub>2</sub> incubator at 37 °C. The dose-response curve of the inhibitor displayed 11 concentrations resulting from serial two-fold dilutions (1:1). The starting concentration of compound A was 100nM. One control well (without any compound) was also included in the course of the experiment. The 24 well plates were incubated for 4 days in a 5% CO<sub>2</sub> incubator at 37 °C. Following a 4 day incubation period, the cells were washed once with PBS and RNA was extracted with the RNeasy® Mini Kit and Qiashredder® from Qiagen. RNA from each well was eluted in 50uL of H<sub>2</sub>O. The RNA was quantified by optical density at 260nm on a Cary 1E UV-Visible Spectrophotometer. 50 ng of RNA from each well was used to quantify the HCV replicon RNA copy number as detailed in Example 6. The level of inhibition (% inhibition) of each well containing inhibitor was calculated with the following

PCT/CA01/01843

equation (CN = HCV Replicon copy number):

$$\% \cdot inhibition = \left(\frac{CN \cdot control - CN \cdot well}{CN \cdot control}\right) * 100$$

5 The calculated % inhibition values were then used to determine IC<sub>50</sub>, slope factor (n) and maximum inhibition (I<sub>max</sub>) by the non-linear regression routine NLIN procedure of SAS using the following equation:

$$\% \cdot inhibition = \frac{I_{\text{max}} \times [inhibitor]^n}{[inhibitor]^n + IC_{50}^n}$$

10

Compound A was tested in the assay at least 4 times. The IC $_{50}$  curves were analyzed individually by the SAS nonlinear regression analysis. Figure 8 shows a typical curve and Table 2 shows the individual and average IC $_{50}$  values of compound A. The average IC $_{50}$  of compound A in the replication assay was 1.1nM.

15

TABLE 2

IC<sub>50</sub> of compound A in the S22.3 Cell line Replicon Assay.

Compound	IC <sub>50</sub> (nM)	Average IC <sub>50</sub> (nM)
	1.2	•
Α	1.2	,
, ,	1.0	
	0.9	
	1.1 ± 0.2	

# 20 DISCUSSION

25

The reproducible and robust *ex vivo* propagation of hepatitis C virus, to levels required for the accurate testing of potential anti-viral compounds, has not been achieved with any system. As an alternative approach to studying the molecular mechanisms of hepatitis C virus RNA replication, selectable self-replicating bicistronic RNAs were developed (Lohman *et al.*, 1999, Science 285:110-113; Bartenschlager CA 2,303,526). Minimally, these replicons encode for some or all of

28

the non-structural proteins and also carry a selectable marker such as the neomycin phosphotransferase. Though intracellular steady-state levels of these sub-genomic replicon RNAs among the selected clones is moderate to high, the frequency of generating G418-resistant colonies upon transfection of the consensus RNA described by Lohman et al. or Bartenschlager is very low. Less than 100 colonies are generated when 8 million cells are transfected with 1 µg of in vitro transcribed bicistronic replicon RNA. A low efficiency of colony formation was first noted by Lohmann et al (1999 et al, Science 285:110-113). Since then, Lohmann et al. (2001), Blight et al. (2000), and Guo et al. (2001), have isolated sub-genomic RNAs with markedly improved efficiencies in the colony formation assay. Lohmann et al., 1999 originally reported that selection of sub genomic replicons may not involve the selection of adaptive mutants as serially passaged RNA did not demonstrate an improved transfection efficiency. Nevertheless, in an effort to characterize the function and fitness of replicating HCV RNA, we serially passaged the replicon RNA that was isolated from the first selected cell-line. Notably, a significant increase in colony forming efficiency was obtained from this experiment, even though the quantity of replicon RNA was orders of magnitude lower than originally used to transfect the in vitro transcribed RNA. Furthermore, a second round serial passage of replicon RNA from this first generation clone into naive Huh-7 cells provided for yet another increase in colony formation efficiency (Table 1).

Our analysis of replicating HCV RNAs identified several adaptive mutations that enhance the efficiency of colony formation by up to 4 orders of magnitude. Adaptive mutations were found in many non-structural proteins, as well as in the 5' non-translated region. The substitution of the 5'-GG doublet for a 5'-A as the inaugurating nucleotide of the HCV 5'-UTR is a variant of the HCV genome that has not been previously described, despite the sequencing of innumerable genotypes and subtypes from across the world. Our original replicon that carried a 5'-GG evolved to variants with either a single 5'-A or 5'-G, both of which showed equal transduction efficiency. We describe here the first report of a HCV genome that can tolerate and stably maintain a 5'A extremity. Moreover, we were successful in re-introducing this defined single nucleotide substitution into our cDNA clone and generate *in vitro* transcribed RNA harboring such an extremity to confirm that a 5'A functions as efficiently as a 5'G.

10

15

20

25

30

We have identified adaptive amino acid substitutions in the HCV non-structural proteins NS3, NS4A and NS5A in the R3 replicon, and a substitution in NS5B in the R7 clone (see Figure 5B). These mutations, particularly the combination defined by the R3-rep (SEQ ID NO. 7), when reconstituted into a cDNA clone and transcribed onto a RNA replicon, result in a significantly enhanced transduction efficiency of up to 20,000 fold from the original wild type APGK12 replicon RNA. However, the steady state levels of intracellular replicon RNA were comparable from each of the different isolated clones. This result suggests that the increase in replication efficiency by the adaptive mutations does not result in higher stable intracellular RNA levels due to higher RNA replication, but rather confers increased permissivity for establishing the replicon in a greater number of Huh7 cells. Such a phenotype may be manifested transiently, through an initial increase of the amount of *de novo* replication, that is required to surpass a defined threshold to establish persistently replicating RNAs within a population of dividing cells.

Recently three other groups also identified other distinct adaptive mutants. Lohmann *et al.* (2000) reported enhanced transduction efficiencies of up to 10,000 fold with mutations in NS3, NS4B, NS5A and NS5B. Blight *et al.* (2000) reported an augmentation of transduction efficiencies up to 20,000 fold with a single mutation in NS5A whereas Guo *et al.* (2001) reported increases in transduction efficiencies of 5,000-10,000 fold with a deletion of a single amino acid in NS5A. The amino acid substitutions that we describe here have not previously been identified as adaptive mutants that enhance the efficiency of RNA transfection and/or replication. One exception is the mutation of E1202G in NS3 that we found in both the R7 and R16 replicons. This adaptation was previously described by Guo et al (2001) and Krieger et al (2001). All other adaptive mutations, without exception, described herein are unpublished.

The development of selectable subgenomic HCV replicons has provided for potential avenues of exploration on HCV RNA replication, persistence, and pathogenesis in cultured cells. However, the low transduction efficiency with the HCV RNA-containing replicons as originally described (Lohmann et al., 1999) showed that it was not a practical system for reverse genetics studies. The adaptive mutants described herein overcome the low transduction efficiency. In light of the recent descriptions of adaptive mutants by other groups, we note that adaptation can be

PCT/CA01/01843

achieved by distinct mutations in different HCV NS proteins, although the level of adaptation can vary drastically. The replicons encoding adaptive mutants that are described herein are ideally suited for reverse genetic studies to identify novel HCV targets or host cell targets that may modulate HCV RNA replication or HCV replicon RNA colony formation. The adapted and highly efficient replicons are suitable tools for characterizing subtle genotypic or phenotypic changes that affect an easily quantifiable transduction efficiency.

Lastly, we have used our adapted HCV sub genomic replicon cell-line to

demonstrate the proficient inhibition of HCV RNA replication by a specific small molecule inhibitor of the HCV NS3 protease. This is the first demonstration that an antiviral, designed to specifically inhibit one of the HCV non-structural proteins, inhibits HCV RNA replication in cell culture. Moreover, this compound and our S22.3 cell line validate the proposal that RNA replication is directed by the HCV non-structural proteins NS3 to NS5B. The assay that we have described and validated will be extremely useful in characterizing other inhibitors of HCV non-structural protein function in cell culture in a high throughput fashion.

All references found throughout the present disclosure are herein incorporated by reference whether they be found in the following list or not.

# References

WO 02/052015

Ago et al. 1999, Structure 7: 1417-1426

Ausubel et al., 1994, Current Protocols in Molecular Biology, Wiley, New York.

25 Bartenschlager, R. et al., 1993, J. Virol., 67, 3835-3844.

Behrens et al., 1998, J. Virol. 72, 2364

Bressanelli et al. 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039

Blight et al. 2000, Science 290: 1972-1974

Cho et al., 1998, J. Biol. Chem., 273, 15045

30 Dash et al., 1997, Am. J. Pathol., 151, 363 – 373

Fourner et al. 1998, J. Gen. Virol. 79, 2376

Gale Jr. et al. 1997 Virology 230, 217

Grakoui, A. et al., 1993(a), J. Virol. 67, 1385-1395.

Grakoui A, et al., 1993(b), Proc Natl Acad Sci USA, 90, 10583-7

35 Guo et al. (2001) J. Virol. 8516-8523

Hijikata, M. et al., 1991, Proc. Natl. Acad. Sci. USA. 88, 5547-5551.

Hijikata, M. et al., 1993, J. Virol. 67, 4665-4675.

Ikda et al. 1998, Virus Res. 56, 157

Ito et al. 1996, J. Gen. Virol. 77, 1043 - 1054

5 IUPAC-IUB Biochemical Nomenclature Commission, 1972, Biochemistry, 11, 1726-1732.

Kolykhalov et al. 1996 J. of Virology, 7, p. 3363-3371

Khromykh et al., 1997, J. Virol. 71, 1497

Kim, D.W. et al., 1995, Biochem. Biophys. Res. Comm., 215, 160-166.

10 Kim et al., **1996**, Cell, 87, 343;

Kim et al., 1998, Structure, 6, 89

Kim et al., 1999, Arch. Virol, 144, 329-343.

Krieger et al. (2001) J. Virol. 4614-4624

Kwong AD. et al., 1998, Antiviral Res., 40, 1-18

15 Lanford et al. 1994, Virology 202, 606

Lesburg et al. 1999, Nat. Struct. Biol. 6: 937-943

Lohman et al. 1999, Science 285: 110-113

Lohman et al. (2001) J. Virol. 1437-1449

Love, R.A. et al., 1996, Cell, 87, 331-342

20 Martell et al. 1999 J. Clin. Microbiol. 37: 327-332

Mizutani et al. 1996, J. Virol. 70, 7219 – 7223

Moser et al., 1998, J. Virol. 72, 5318

Reed et al., 1997, J. Virol. 71, 7187

Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, Cold Spring

25 Harbor Labs.

Shimizu et al. 1993, PNAS, USA, 90, 6037 - 6041

Yanagi et al., 1999, Proc. Natl. Acad. Sci. USA, 96, 2291-95

Yao et al., 1997, Nature Structural Biology, 4, 463

Yem et al., 1998, Protein Science, 7, 837

30 Yoo et al. **1995**, J. Virol., 69, 32 – 38

32

# CLAIMS

- 1. A HCV polynucleotide molecule comprising a 5'-non translated region (NTR) wherein guanine at position 1 is substituted for adenine.
- 2. A HCV self-replicating polynucleotide comprising:
  - a 5'-NTR consisting of ACCAGC (SEQ ID NO. 8);
  - a HCV polyprotein region coding for a HCV polyprotein; and
  - a 3'-NTR region.
- The HCV polynucleotide according to claim 2, wherein said polyprotein comprises one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.
- 4. The HCV polynucleotide encoding a polyprotein comprising one or more of the amino acid substitution as defined in claim 3, and further comprising the amino acid substitution E(1202)G.
- 5. The HCV polynucleotide according to claim 3, wherein said substitution is a G2042C or a G2042R mutation.
- 6. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: K(1691)R; and G(2042)C.
- 7. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 8. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 9. The HCV polynucleotide according to claim 3, wherein said substitution is selected

WO 02/052015

- from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
- 10. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- The HCV polynucleotide according to claim 2, wherein said polynucleotide is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ
   ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 12. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- **13.** An expression vector comprising a DNA form of the polynucleotide according to claim 2, operably linked to a promoter.
- **14.** A host cell transfected with the self-replicating polynucleotide molecule according to claim 2.
- 15. A host cell according to claim 14, wherein the host cell is a eukaryotic cell line.
- **16.** A host cell according to claim 15, wherein said eukaryotic cell line is a hepatic cell line.
- 17. A host cell according to claim 16, wherein said hepatic cell line is Huh-7.
- **18.** A RNA replication assay comprising the steps of:
  - a) incubating the host cell according to claim 14 under conditions suitable for RNA replication;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 19. The assay according to claim 18, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

34

- 20. The assay according to claim 18, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- 21. A method for testing a compound for inhibiting HCV replication, including the steps of:
  - a) carrying step a) according to claim 18, in the presence or absence of the compound;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
  - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.

wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

- 22. The method according to claim 21, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.
- 23. A HCV polynucleotide molecule comprising:
  - a 5'-NTR region;
  - a HCV polyprotein region coding for a HCV polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T; and a 3'-NTR region.
- 24. The HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as defined in claim 24, further comprising the amino acid substitution E(1202)G.
- **25.** The polynucleotide according to claim 24, wherein said substitution is a G2042C or a G2042R mutation.
- 26. The HCV polynucleotide according to claim 24, wherein said substitution is selected

- from the group consisting of: K(1691)R; and G(2042)C.
- 27. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 28. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 29. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
- 30. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- 31. The HCV polynucleotide according to claim 24, wherein said molecule is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 32. The HCV polynucleotide according to claim 24, wherein said molecule is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 33. An expression vector comprising a DNA form of the polynucleotide according to claim 24, operably linked to a promoter.
- 34. A host cell transfected with the self-replicating polynucleotide according to claim 24.
- 35. A host cell according to claim 34, wherein the host cell is a eukaryotic cell line.
- 36. A host cell according to claim 35, wherein said eukaryotic cell line is a hepatic cell line.
- 37. A host cell according to claim 36, wherein said hepatic cell line is Huh-7.

38. A RNA replication assay comprising the steps of:

incubating the host cell according to claim 34 under conditions suitable for RNA replication;

isolating the total cellular RNA from the cells; and analyzing the RNA so as to measure the amount of HCV RNA replicated.

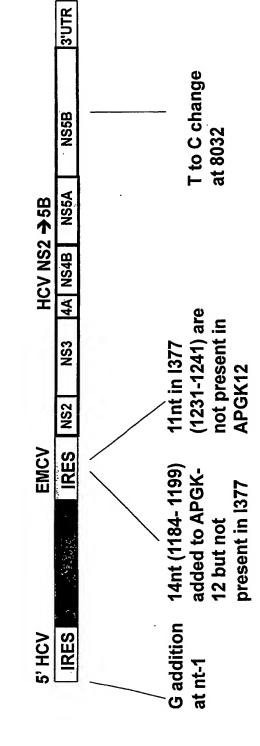
- 39. The assay according to claim 38, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.
- 40. The assay according to claim 38, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- **41.** A method for testing a compound for inhibiting HCV replication, including the steps of:
  - a) carrying step a) according to claim 38, in the presence or absence of the compound;
  - b) isolating the total cellular RNA from the cells; and
  - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
  - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.

wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

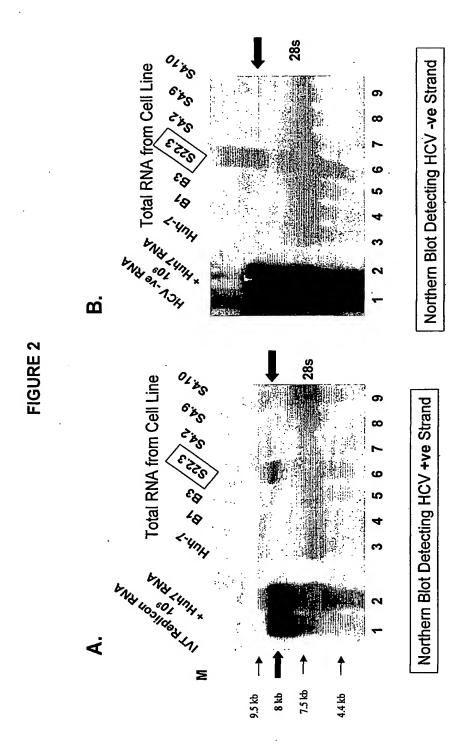
**42.** The method according to claim 41, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

FIGURE 1

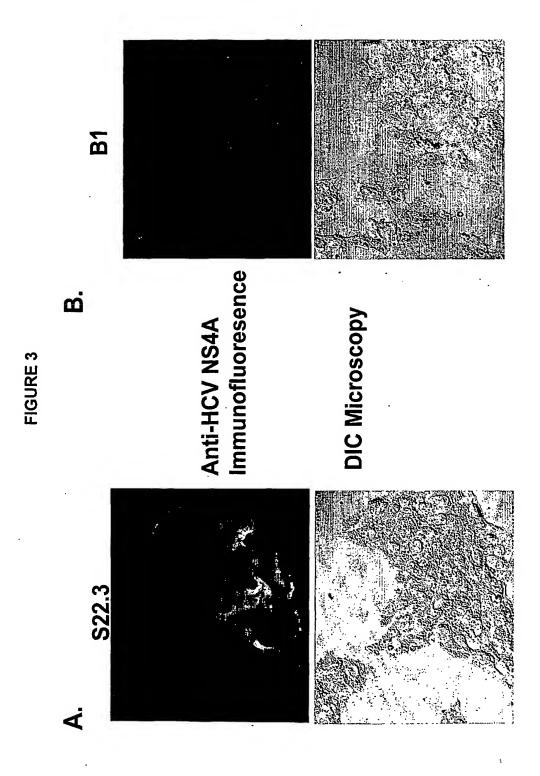
APGK12 (SEQ ID NO 1) Replicon RNA compared to I377/NS2-3' Replicon RNA



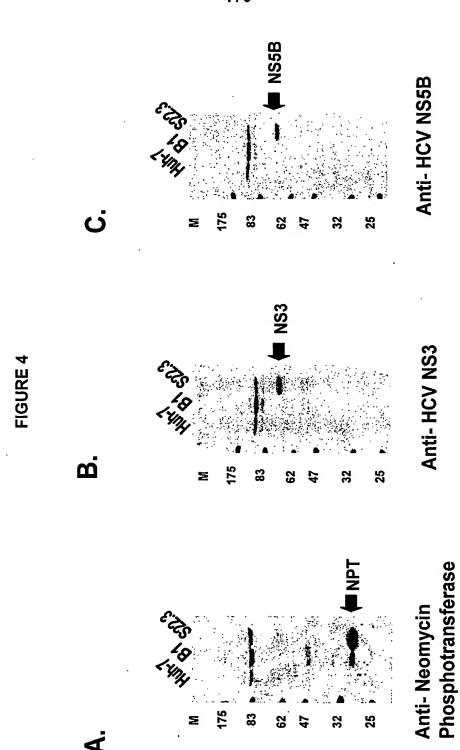
**SUBSTITUTE SHEET (RULE 26)** 



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

5/9

**IGURE 5A** 

	S 22-3 SEQ ID	R3 SEQ (D	R3-rep SEQ ID	R7 SEQ ID	R16 SEQ ID
	NO 2	NO. 4	NO. 7	NO. 5	9 ON
'end - FIRST nt (HCV IRES)	*G (nt 1) A	G (nt 1) A		•	G (nt 1) A
Neo	. •	A (nt 461) G	•		
EMCV IRES	•	A (nt 1739) G	•		
NS 2		•		· .	
. NS 3		G (nt 2778) A A (nt 2840) C A (nt 4052) G	T (nt 2509) C G (nt 2778) A A (nt 2840) C T (nt 3574) C A (nt 4052) G	A (nt 2936) G A (nt 2979) G	A (nt 2816) G A (nt 2979) G
NS 4A	A (nt 4446) R	A (nt 4446) G	C (nt 4387) T A (nt 4446) G C (nt 4507) T	•	C (nt 4475) T
NS 4B	•	T (nt 4855) C	T (nt 4855) C		•
NS 5A	G (nt 5498) T A (nt 6268) R	A (nt 5351) G G (nt 5498) T G (nt 5659) A C (nt 6871) T A (nt 6268) G	A (nt 5361) G G (nt 6498) T G (nt 6669) A T (nt 6838) C C (nt 5871) T A (nt 6116) G	A (nt 5324) G G (nt 5498) T T (nt 6001) C	G (nt 6498) C T (nt 6320) C T (nt 6584) C
NS 5B	•	A ( nt 6652) G	•	C (nt 7252) T T (nt 8349) C	•
3'end - last 98 nt	•	-	:	•	•

\*first nt = G from HCV ires

6/9

-IGURE 5B

	S 22-3 SEQ ID NO. 2	R3 SEQ ID NO. 4	S 22-3 SEQ ID R3 SEQ ID R3 Rep SEQ ID R7 SEQ ID NO. 2 NO. 4 NO. 7 NO. 5	R7 SEQ ID NO. 5	R16 SEQ ID NO. 6
5'end - FIRST nt (HCV IRES)	G (nt 1) A	G (nt 1) A	•	•	G (nt 1) A
NS 2	•	•	•	•	•
NS 3		R (1135) K S (1560) G	R (1135) K S (1560) G	E (1202) G	S (1148) G E (1202) G
NS 4A	K (1691) mix K/R	K (1691) R	K (1691) R		L (1701) F
NS 4B	•	,	•		•
NS 5A	G (2042) C	T (1993) A G (2042) C	T (1993) A G (2042) C	I (1984) V G (2042) C	G (2042) R S (2404) P
		P (2166) L	L (2155) P P (2166) L		
NS 5B	•	•	•	M (2992) T	•
3'end - last 98 nt		•			•

first a.a. of NS2 = 810

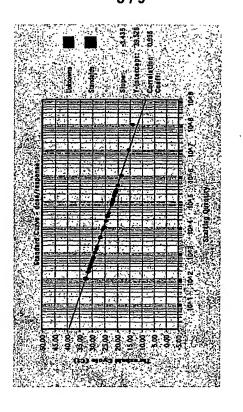
7/9

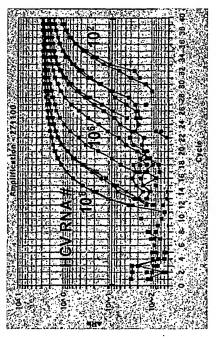
**3'HCV** UTR NS5B T(1993)A G(2042)C L(2155)P L(2155)P G(2042)C P(2166)L T(1993)A P(2166)L NS5A AMINO ACID SUBSTITUTIONS NS4B HCV NS2→5B K(1691)R K(1691)R R(1135)K S(1560)G R(1135)K S(1560)G FIGURE 6 NS3 NS2 EMCV NeoR SEQ ID NO 25 SEQ ID NO 24 SEQ ID NO 1 SEQ ID NO 7 CLONE APGK-12 5' HCV IRES G (nt1) R3 rep A (nt1) A(nt1) G(nt1) 1100000cfu/µg 2000000cfu/µg 77 cfu/µg 86 cfu/µg

**SUBSTITUTE SHEET (RULE 26)** 

HCV-Replicon: RNA Quantification

FIGURE 7

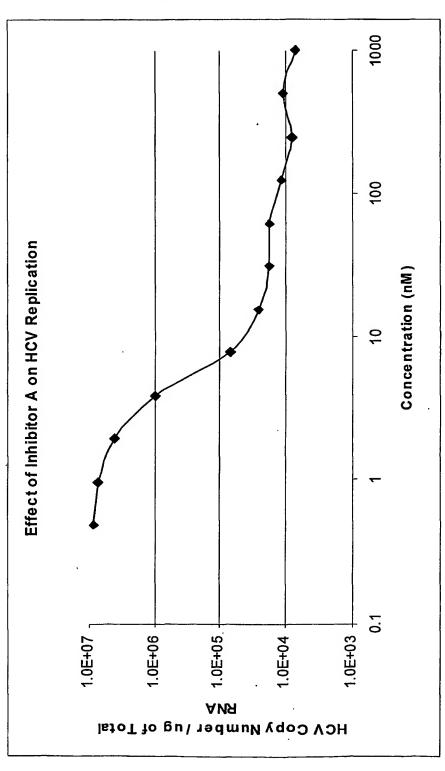




Ct = Threshold cycle  $\alpha$  Starting RNA Quantity

SUBSTITUTE SHEET (RULE 26)





#### 1 / 93

#### **SEQUENCE LISTING**

```
<110> BOEHRINGER INGELHEIM (CANADA) LTD.
<120> SELF REPLICATING RNA MOLECULE FROM
 HEPATITIS C VIRUS
<130> 13/083
<150> 60/257,857
<151> 2000-12-22
<160> 25
<170> FastSEQ for Windows Version 4.0
<210> 1
<211> 8639
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1803)...(8408)
<400> 1
ggccagcccc cgattggggg cgacactcca ccatagatca ctcccctgtg aggaactact 60
gtcttcacgc agaaagcgtc tagccatggc gttagtatga gtgtcgtgca gcctccagga 120
cccccctcc cgggagagcc atagtggtct gcggaaccgg tgagtacacc ggaattgcca 180
ggacgaccgg gtcctttctt ggatcaaccc gctcaatgcc tggagatttg ggcgtgcccc 240
cgcgagactg ctagccgagt agtgttgggt cgcgaaaggc cttgtggtac tgcctgatag 300
ggtgcttgcg agtgccccgg gaggtctcgt agaccgtgca ccatgagcac gaatcctaaa 360
cctcaaagaa aaaccaaagg gcgcgccatg attgaacaag atggattgca cgcaggttct 420
ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac aatcggctgc 480
tetgatgeeg cegtgtteeg getgteageg caggggegee eggttetttt tgteaagace 540
gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc 600
acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg aagggactgg 660
ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc tcctgccgag 720
aaagtateea teatggetga tgeaatgegg eggetgeata egettgatee ggetaeetge 780
ccattegace accaagegaa acategcate gagegageac gtacteggat ggaageeggt 840
cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc 900
gccaggctca aggcgcgcat gcccgacggc gaggatctcg tcgtgaccca tggcgatgcc 960
tgettgeega atateatggt ggaaaatgge egettttetg gatteatega etgtggeegg 1020
ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat tgctgaagag 1080
cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc tcccgattcg 1140
cagegeateg cettetateg cettettgae gagttettet gagttegege ceagatgtta 1200
acagaccaca acggtttccc tctagcggga tcaattccgc ccccccct aacgttactg 1260
gccgaagccg cttggaataa ggccggtgtg cgtttgtcta tatgttattt tccaccatat 1320
tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg acgagcattc 1380
ctaggggtct ttcccctctc gccaaaggaa tgcaaggtct gttgaatgtc gtgaaggaag 1440
cagtteetet ggaagettet tgaagacaaa caacgtetgt agegaceett tgeaggeage 1500
ggaacccccc acctggcgac aggtgcctct geggccaaaa gccacgtgta taagatacac 1560
ctgcaaaggc ggcacaaccc cagtgccacg ttgtgagttg gatagttgtg gaaagagtca 1620
aatggetete eteaagegta tteaacaagg ggetgaagga tgeecagaag gtaccccatt 1680
gtatgggatc tgatctgggg cctcggtgca catgctttac atgtgtttag tcgaggttaa 1740
aaaacgtcta ggccccccga accacgggga cgtggttttc ctttgaaaaa cacgataata 1800
```

											ggc Gly					1847
										Tyr	aag Lys					1895
											agg Arg			Ala		1943
											gly ggg		Arg			1991
		Leu									cta Leu 75					2039
acc Thr 80	aaa Lys	atc Ile	ttg Leu	ctc Leu	gcc Ala 85	ata Ile	ctc Leu	ggt Gly	cca Pro	ctc Leu 90	atg Met	gtg Val	ctc Leu	cag Gln	gct Ala 95	2087
										Ala	cac His					2135
											cat His					2183
-		_	_	-	_	_	_			_	tac Tyr	_	Tyr	_		2231
											cta Leu 155					2279
	_	_			_	_			_	_	gag Glu		_	_		2327
											atc Ile					2375
											ctg Leu					2423
											cct Pro					2471

tcc Ser	caa Gln 225	cag Gln	acg Thr	cga Arg	Gly ggc	cta Leu 230	ctt Leu	gly ggc	tgc Cys	atc Ile	atc Ile 235	act Thr	agc Ser	ctc Leu	aca Thr	2519
	cgg Arg															2567
	aca Thr															2615
	tat Tyr															2663
	acc Thr															2711
	ccc Pro 305				_		_			_		_		_	_	2759
	ctt Leu															2807
	ggc															2855
	aag Lys															2903 ·
	gly															2951
	gac Asp 385		-		_			_	_			_			-	2999
	ttc Phe															3047
	gcc Ala				_				_		_	_		_		3095
	gct Ala															3143

				acc Thr												<b>3191</b>
				aac Asn												3239
_			_	tac Tyr					_			-	_			3287
_				gcc Ala 500		-				_	_		_			3335
				act Thr												3383
				gcg Ala												3431
				gtg Val												3479
_			_	atc Ile												3527
	_			agg Arg 580					-			_	_		_	3575
-				gcg Ala	_	_							_	_	_	3623
				ctt Leu	_	_		_				_		_	_	3671
att Ile	gtc Val 625	gta Val	gca Ala	acg Thr	gac Asp	gct Ala 630	cta Leu	atg Met	acg Thr	Gly	ttt Phe 635	acc Thr	Gly	gat Asp	ttc Phe	3719
				gac Asp												3767
				acc Thr 660												3815

													ggc Gly 685			3863
													ggc Gly			3911
													gct Ala			3959
													tac Tyr			4007
			_		_	_	_	_		_			tgg Trp		_	4055
													tcc Ser 765			4103
													cag Gln			4151
													caa Gln			4199
													acg Thr			4247
													aca Thr			4295
					_	-	_	_	_	_	-	_	gag Glu 845	_	_	4343
acg Thr	agc Ser	acc Thr 850	tgg Trp	gtg Val	ctg Leu	gta Val	ggc 855	gga Gly	gtc Val	cta Leu	gca Ala	gct Ala 860	ctg Leu	gcc Ala	gcg Ala	4391
tat Tyr	tgc Cys 865	ctg Leu	aca Thr	aca Thr	gly ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	4439
tcc Ser 880	gga Gly	aag Lys	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	4487

				gaa Glu 900												4535
	_	_		gcc Ala	_				_	_	_			_	_	4583
				aag Lys												4631
				ctc Leu												4679
				caa Gln												4727
			-	tca Ser 980	-	_	-			-				-	-	4775
				cat His 5					Asn					$\mathtt{Trp}$		4823
			Leu	gct Ala				Ala					Val			4871
		Ala		gcg Ala			Gly					Gly				4919
	Asp			gca Ala		Tyr					Ala					4967
_		_	_	atg Met 1060	Ser			_		Ser		_	Asp	_	Val	5015
				gct Ala 5					Gly					Gly		5063
			Ala	ata Ile				His					Glu			5111
		Trp		aac Asn			Ile					Arg				5159

		cac tat His Tyr 1125	Val Pro					5207
		tct agt Ser Ser 1140			Gln Leu			5255
		aac gag Asn Glu 5					Ser Trp	5303
_	-	tgg gat Trp Asp		Cys Thr		_	_	5351
	Leu Gln	tcc aag Ser Lys				Gly Val		5399
	-	cgt ggg Arg Gly 1205	Tyr Lys					5447
_		tgc cca Cys Pro 1220			lle Thr			5495
	_	agg atc Arg Ile 5				-	Thr Trp	5543
		ccc att Pro Ile		Tyr Thr				5591
tcc ccg Ser Pro 1265	Ala Pro	aat tat Asn Tyr	tct agg Ser Arg 1270	gcg ctg Ala Leu	tgg cgg Trp Arg 127	Val Ala	gct gag Ala Glu	5639
		gtt acg Val Thr 1285	Arg Val					5687
		aac gta Asn Val 1300			Gln Val			5735
		gtg gat Val Asp 5					Pro Ala	5783
tgc aaa Cys Lys	ccc ctc Pro Leu 1330	cta cgg Leu Arg	gag gag Glu Glu 133	Val Thr	ttc ctg Phe Leu	gtc ggg Val Gly 1340	ctc aat Leu Asn	5831

		Leu					Leu					Glu	ccg Pro			5879
	Val					Leu					His		acg Thr			5927
					Leu					${\tt Pro}$			ttg Leu		Ser	5975
		_	_	Gln	_				Ser	_	_	_	aca Thr 140	Cys		6023
			Asp					Asp					aac Asn O			6071
		Gln					Asn					Glu	tca Ser			6119
	Val					Ser					Gln		gag Glu			6167
					Val					Leu			tcc Ser		ГÀЗ	6215
				Met					Arg				aac Asn 148!	Pro		6263
_			Ser		_	_	_	Asp		_			gtg Val )	_		6311
		Pro					Lys					Pro	cct Pro			6359
	Lys		_	_	_	Leu		_			Val		tct Ser	_	_	6407
					Lys					Ser			tcg Ser		Val	6455
				Ala					Asp				gac Asp 156	Asp		6503

Asp	gcg Ala		Ser					Tyr					Pro			6551
	gag Glu 1589	Pro					Leu					Trp				6599
	gag Glu O					qaA					Ser					6647
	aca Thr		_	_	Ile	. –		_	_	Ala		_		-	Leu	6695
	atc Ile			Leu					Leu					Leu		6743
	gct Ala		Thr		_	-	-	Ser	_		_	_	Lys	_		6791
	gac Asp 1669	Arg					Asp					Asp				6839
- дад																
	Met					Ser				gct Ala 1690	Lys					6887
Glu 1680 gag	Met	Lya	Ala	Lys	Ala 1685 ctg Leu	Ser acg	Thr	Val	Lys	Ala 1690 tcg Ser	gcc Lys	Leu aga	Leu.	Ser , aaa	Val 1695 ttt Phe	6935
Glu 1680 gag Glu ggc	Met ) gaa Glu	gcc Ala	Ala tgt Cys gca	Lys aag Lys 1700 aag Lys	Ala 1689 ctg Leu )	Ser acg Thr	Thr ccc Pro	Val cca Pro	cat His 1705 cta Leu	Ala 1690 tcg Ser	Lys gcc Ala agc	Leu aga Arg aag	tct Ser	ser aaa Lys 1710 gtt Val	Val 1695 ttt Phe	
Glu 1680 gag Glu ggc Gly cac	Met gaa Glu tat	gcc Ala ggg Gly	tgt Cys gca Ala 1715 tcc Ser	Lys aag Lys 1700 aag Lys Gys	Ala 1685 ctg Leu gac Asp	ser acg Thr gtc Val	Thr ccc Pro cgg Arg	Val cca Pro aac Asn 1720 ttg Leu	cat His 1705 cta Leu	tcg Ser tcc Ser	Lys ) gcc Ala agc ser	Leu aga Arg aag Lys	tct Ser gcc Ala 1725 gag Glu	aaa Lys 1710 gtt Val	Val 1695 ttt Phe ) aac Asn	6935
Glu 1680 gag Glu ggc Gly cac His	Met ) gaa Glu tat Tyr	gcc Ala ggg Gly cgc Arg 1730 acc Thr	tgt Cys gca Ala 1715 tcc Ser	aag Lys 1700 aag Lys gtg Val	Ala 1685 ctg Leu gac Asp tgg Trp	ser acg Thr gtc Val aag Lys gca	Thr  ccc Pro  cgg Arg  gac Asp 1735 aaa Lys	Val  cca Pro  aac Asn 1720  ttg Leu  aat	cat His 1705 cta Leu ctg Leu	tcg ser tcc ser gaa Glu	gcc Ala agc ser gac Asp	aga Arg aag Lys act Thr 1740 tgc Cys	tct Ser gcc Ala 1725 gag Glu	aaa Lys 1710 gtt Val aca Thr	Val 1695 ttt Phe aac Asn	6935 6983
gag Glu ggc Gly cac His att Ile	gaa Glu tat Tyr atc Ile gac Asp 1745 aag Lys	gcc Ala ggg Gly cgc Arg 1730 acc Thr	tgt Cys gca Ala 1715 tcc Ser acc Thr	aag Lys 1700 aag Lys 5 gtg Val atc Ile	ctg Leu gac Asp tgg Trp	acg Thr gtc Val aag Lys gca Ala 1750	Thr  CCC Pro  CGG Arg  Gac Asp 1735 aaa Lys	CCA Pro aac Asn 1720 ttg Leu aat Asn	cat His 1705 cta Leu ctg Leu gag Glu	tcg ser tcc ser gaa Glu gtt Val	gcc Ala agc ser gac Asp ttc Phe 1755 gta Val	aga Arg aag Lys act Thr 1740 tgc Cys	tct Ser gcc Ala 1725 gag Glu gtc Val	ser . aaa Lys 1710 gtt Val . aca Thr	Val 1695 ttt Phe aac Asn cca Pro	6935 6983 7031

# .10 / 93

ctc cct cag gcc Leu Pro Gln Ala 1795	Val Met Gly				
gga cag cgg gtc g Gly Gln Arg Val				Lys Lys Cy	
cct atg ggc ttc g Pro Met Gly Phe . 1825		Thr Arg Cys			
gag aat gac atc Glu Asn Asp Ile : 1840				Cys Asp Le	
gcc ccc gaa gcc Ala Pro Glu Ala			Leu Thr Glu		
atc ggg ggc ccc Ile Gly Gly Pro 1 1875					
cgg tgc cgc gcg a Arg Cys Arg Ala ( 1890				Asn Thr Le	
aca tgt tac ttg a Thr Cys Tyr Leu : 1905		Ala Ala Cys			
gac tgc acg atg of Asp Cys Thr Met 1				Ile Cys Gl	
age geg ggg ace of Ser Ala Gly Thr			Leu Arg Ala		
gct atg act aga Ala Met Thr Arg 1955					
tac gac ttg gag ( Tyr Asp Leu Glu 1 1970	ttg ata aca Leu Ile Thr	tca tgc tcc Ser Cys Ser 1975	tcc aat gtg Ser Asn Val 1980	Ser Val Al	g 7751 a
cac gat gca tct q His Asp Ala Ser ( 1985		Val Tyr Tyr			
acc ccc ctt gcg of Thr Pro Leu Ala 2 2000					1

#### 11 / 93

		y Asn Ile				tg tgg gca Leu Trp Ala 2030	7895
				Ser Ile	Leu Leu A	gct cag gaa Ala Gln Glu 2045	7943
Gln Leu G		_				gt tac tcc Cys Tyr Ser	7991
			Gln Ile			cac ggc ctt His Gly Leu	8039
					Glu Ile A	aat agg gtg Asn Arg Val 2095	8087
gct tca t Ala Ser C	gc ctc ag Cys Léu Ar 21	g Lys Leu	ggg gta Gly Val	ccg ccc Pro Pro 2105	ttg cga g Leu Arg V	gtc tgg aga Val Trp Arg 2110	8135
				Leu Leu	Ser Gln G	ggg ggg agg Sly Gly Arg 2125	8183
Ala Ala T						ngg acc aag Arg Thr Lys	8231
		_	Ala Ala	_		ta tcc agc Leu Ser Ser	8279
					Tyr His S	agc ctg tct Ser Leu Ser 2175	8327
	-	g Trp Phe		_		ctt tct gta Leu Ser Val 2190	8375
	ggc atc ta ly Ile Ty 2195			Arg *	acggggagc	ct aaacactcca	8428
tttttttt ccatcttag	g ccatcct	ctt ttttt acg gctago	ccctt tt tcct ct ctgtg aaa	ttttttc tttttcc	ttttctttc	t ttttttttt c tttggtggct g actgcagaga	8548

<210> 2 <211> 8642 <212> DNA <213> HCV

<220>

```
<221> CDS
<222> (1802)...(8407)
<221> variation
<222> 6268
<223> r = a or g
<221> variation
<222> 4446
<223> r = a or q
accageccce gattggggge gacactecae catagateae teecetgtga ggaactaetg 60
tetteaegea gaaagegtet agceatggeg ttagtatgag tgtegtgeag ceteeaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaceggg teetttettg gateaacecg etcaatgeet ggagatttgg gegtgeeece 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagegc aggggcgcc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcage gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatac gettgatecg getacetgee 780
cattegacca ccaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggeteaa ggegegeatg ceegaeggeg aggatetegt egtgaeeeat ggegatgeet 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggetgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee cagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt teceeteteg ecaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440
agtteetetg gaagettett gaagacaaac aacgtetgta gegaceettt geaggeageg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctetec tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtetag geceeegaa eeacggggae gtggttttee tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg
                                                                  1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                  1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                  1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
    50
                                             60
                         55
```

# 13/93 .

	ctc Leu		_	_												2041
	atc Ile															2089
	acc Thr															2137
	atg Met															2185
	atg Met 130	_	_	_	_	_			_		_		_			2233
	cca Pro															2281
	gtt Val														acc Thr	2329
	gjà aaa															2377
	tcc Ser															2425
	gaa Glu 210		_			_						_	_			2473
	cag Gln	_	_					_				_				2521
	gac Asp															2569
aca Thr	caa Gln	tct Ser	ttc Phe 260	ctg Leu	gcg Ala	acc Thr	tgc Cys	gtc Val 265	aat Asn	ggc	gtg Val	tgt Cys	tgg Trp 270	act Thr	gtc Val	2617
tat Tyr	cat His	ggt Gly 275	gcc Ala	ggc Gly	tca Ser	aag Lys	acc Thr 280	ctt Leu	gcc Ala	ggc Gly	cca Pro	aag Lys 285	ggc	cca Pro	atc Ile	2665

								gtc Val 300					2713
								tgc Cys					2761
								ccg Pro					2809
								ccc Pro					2857
								tcg Ser					2905
								gtt Val 380					2953
								atg Met					3001
								cag Gln					3049
-			-		 _		_	agc Ser		_		_	3097
_		_	-		_	-		gtc Val	_		_		3145
								tct Ser 460					3193
								atc Ile					3241
								gcc Ala					3289
								gag Glu					3337

gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	ggc ggc	atc Ile 520	ggc Gly	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gln	gcg Ala	gag Glu	3385
											gct Ala 540					3433
	-										gtg Val					3481
		_							-		ccc Pro					3529
											aag Lys					3577
		_		_							aat Asn	_	_	-		3625
											agc Ser 620					3673
~	_	~	_	_	_		_				acc Thr	~~	_		_	3721
											aca Thr					3769
											gtg Val					3817
											agg Arg					3865
								_			tcg Ser 700		_		_	3913
											tgt Cys					3961
											gct Ala					4009

											ttc Phe					4057
	Thr						_	_			ttg Leu		_		_	4105
											tac Tyr 780					4153
-	_		_	_	_				_		gac Asp		_		_	4201
											cca Pro					4249
											acc Thr					4297
											ctg Leu					4345
											gct Ala 860					4393
											agg Arg					4441
											ctt Leu					4489
											tac Tyr					4537
											atc Ile					4585
aca Thr	gcc Ala 930	acc Thr	aag Lys	caa Gln	gcg Ala	gag Glu 935	gct Ala	gct Ala	gct Ala	ccc Pro	gtg Val 940	gtg Val	gaa Glu	tcc Ser	aag Lys	4633
tgg Trp	cgg	acc	ctc	gaa	gcc	ttq	tgg	gcg	aag	cat	atg	tgg	aat	ttc	atc	4681

		ata Ile														4729
		gca Ala														4777
		caa Gln 999	His			_		Asn		_			Trp		_	4825
_		ctt Leu	_			_	Ala	_		_		Val		_		4873
	Ala	gga Gly		_	_	Gly	_				Gly	_				4921
		ttg Leu			$\mathtt{Tyr}$					Ala					Ala	4969
		gtc Val		Ser					Ser					Val		5017
cta Leu	ctc Leu	cct Pro 1075	Ala	atc Ile	ctc Leu	tcc Ser	cct Pro 1080	Gly	gcc Ala	cta Leu	gtc Val	gtc Val 1089	${\tt Gly}$	gtc Val	gtg Val	5065
		gcg Ala O					His					Glu				5113
_		atg	aac	caa												
	_	Met	Asn		_	Ile			gct Ala	_	Arg				_	5161
tcc	cac	Met acg Thr	cac	Arg	Leu 1110 gtg Val	Ile ) cct	Ala	Phe agc	Ala	Ser 1115 gct Ala	Arg gca	Gly	Asn cgt	His gtc	Val 1120 act Thr	5161 5209
tcc Ser	ccc Pro	acg	cac His	Arg tat Tyr 1125 agt Ser	Leu 1110 gtg Val	cct Pro	Ala gag Glu atc	Phe agc Ser	gac Asp 1130 cag Gln	Ser 1115 gct Ala )	Arg gca Ala	Gly gca Ala aag	Asn cgt Arg	gtc Val 1135 ctt Leu	Val 1120 act Thr	
tcc Ser cag Gln	ccc Pro atc Ile	acg Thr	cac His tct Ser 1140	tat Tyr 1125 agt Ser	Leu 1110 gtg Val ctt Leu gac	Ile  cct Pro  acc Thr	Ala gag Glu atc Ile	agc ser act Thr 1145 acg	gac Asp 1130 cag Gln	gct Ala ) ctg Leu	gca Ala ctg Leu	Gly gca Ala aag Lys	agg Arg 1150 tcg Ser	gtc Val 1135 ctt Leu	Val 1120 act Thr cac His	5209

Leu			aag Lys		Leu					Glу					5401
			999 Gly 120	Tyr					Arg					Met	5449
			cca Pro					Ile					Lys		5497
		Arg	atc Ile				Arg					Thr			5545
	Phe		att			Tyr		_			Cys	_			5593
Ala			tat Tyr		Arg					Val					5641
			acg Thr 128	Arg					His					Met	5689
			gta Val O					Gln					Glu		5737
		Val	gat Asp				Leu					Pro			5785
	Leu		cgg Arg			Val					Gly				5833
Leu			tca Ser		Leu					Glu					5881
			atg Met 1369	Leu					His					Thr	5929
			ctg Leu )					Pro					Ser		5977
		Gln	ctg Leu				Ser					Cys			6025

cgt cat gac tcc Arg His Asp Ser 1410				
cgg cag gag atg Arg Gln Glu Met 1425			Glu Ser Glu	
gta gta att ttg Val Val Ile Leu				
agg gaa gta tcc Arg Glu Val Ser 146	Val Pro Ala G			Lys Phe
cct cga gcg atg Pro Arg Ala Met 1475	Pro Ile Trp A			
ttr gag tcc tgg Xaa Glu Ser Trp 1490		_		
tgt cca ttg ccg Cys Pro Leu Pro 1505		_	Pro Pro Pro	
aag agg acg gtt Lys Arg Thr Val				
gag ctc gcc aca Glu Leu Ala Thr 154	Lys Thr Phe G			Val Asp
agc ggc acg gca Ser Gly Thr Ala 1555	Thr Ala Ser P			
gcg gga tcc gac Ala Gly Ser Asp 1570				
gag ccg ggg gat Glu Pro Gly Asp 1585			Trp Ser Thr	
gag gag gct agt Glu Glu Ala Ser				
aca ggc gcc ctg Thr Gly Ala Leu . 162	atc acg cca to	ge get. geg gag		ctg ccc 6697 Leu Pro
atc aat gca ctg Ile Asn Ala Leu 1635	Ser Asn Ser Le			

1650	er Arg Ser A		gg cag aag aag rg Gln Lys Lys 1660	
gac aga ctg ca Asp Arg Leu Gl 1665	ag gtc ctg g In Val Leu A 1670	gac gac cac to Asp Asp His T	ac cgg gac gtg yr Arg Asp Val 1675	ctc aag gag 6841 Leu Lys Glu 1680
		Thr Val Lys A	ct aaa ctt cta la Lys Leu Leu 690	
Glu Ala Cys Ly			cg gcc aga tct er Ala Arg Ser	
			cc agc aag gcc er Ser Lys Ala 1729	Val Asn His
	al Trp Lys A		aa gac act gag lu Asp Thr Glu 1740	
_			tt ttc tgc gtc al Phe Cys Val 1755	
0 000 00 0		get ege ett a	tc gta ttc cca	gat ttg ggg 7129
Lys Gly Gly Ar	rg Lys Pro A 1765		le Val Phe Pro 770	Asp Leu Gly 1775
gtt cgt gtg tg Val Arg Val Cy	1765 gc gag aaa a	1 stg gcc ctt ta		1775' tec acc etc 7177
gtt cgt gtg tg Val Arg Val Cy 17	1765 gc gag aaa a ys Glu Lys M 780 cg atg ggc t	atg gcc ctt to Met Ala Leu T 1785 cct tca tac g	770 ac gat gtg gtc	tcc acc ctc 7177 Ser Thr Leu 1790 tct cct gga 7225 Ser Pro Gly
gtt cgt gtg tg Val Arg Val Cy 17 cct cag gcc gt Pro Gln Ala Va 1795 cag cgg gtc ga	1765 gc gag aaa a ys Glu Lys M 780 gg atg ggc t al Met Gly s ag ttc ctg g lu Phe Leu V	atg gcc ctt to Met Ala Leu T 1785 cct tca tac gg Ser Ser Tyr G 1800 gtg aat gcc t	ac gat gtg gtc yr Asp Val Val ga ttc caa tac ly Phe Gln Tyr	tcc acc ctc 7177 Ser Thr Leu 1790  tct cct gga 7225 Ser Pro Gly  aaa tgc cct 7273
gtt cgt gtg tg Val Arg Val Cy 17 cct cag gcc gt Pro Gln Ala Va 1795 cag cgg gtc ga Gln Arg Val Gl 1810 atg ggc ttc gc	1765  gc gag aaa a ys Glu Lys M 780  cg atg ggc t al Met Gly S  . ag ttc ctg g lu Phe Leu V 1 ca tat gac a	atg gcc ctt ta Met Ala Leu T 1785  cct tca tac gg Ser Ser Tyr G 1800  gtg aat gcc t Val Asn Ala T 1815	ac gat gtg gtc yr Asp Val Val  ga ttc caa tac ly Phe Gln Tyr 1809 gg aaa gcg aag rp Lys Ala Lys	tcc acc ctc 7177 Ser Thr Leu 1790  tct cct gga 7225 Ser Pro Gly  aaa tgc cct 7273 Lys Cys Pro  gtc act gag 7321
gtt cgt gtg tg Val Arg Val Cy 17 cct cag gcc gt Pro Gln Ala Va 1795 cag cgg gtc ga Gln Arg Val Gl 1810 atg ggc ttc gc Met Gly Phe Al 1825 aat gac atc cg	1765 gc gag aaa a ys Glu Lys M 780 cg atg ggc t al Met Gly S ag ttc ctg g lu Phe Leu V 1 ca tat gac a la Tyr Asp T 1830 gt gtt gag g	atg gcc ctt to Met Ala Leu T 1785  cct tca tac gg Ger Ser Tyr G 1800  gtg aat gcc te Wal Asn Ala T 1815  acc cgc tgt to Thr Arg Cys P 1810  gag tca atc to Slu Ser Ile T	ac gat gtg gtc yr Asp Val Val  ga ttc caa tac ly Phe Gln Tyr 1809 gg aaa gcg aag rp Lys Ala Lys 1820  tt gac tca acg he Asp Ser Thr	tcc acc ctc 7177 Ser Thr Leu 1790  tct cct gga 7225 Ser Pro Gly  aaa tgc cct 7273 Lys Cys Pro  gtc act gag 7321 Val Thr Glu 1840  gac ttg gcc 7369

			Leu			tct Ser		Gly					Tyr			7465
		Ala				ctg Leu 1899	Thr					Asn				7513
	Tyr					gcg Ala ) ·					Ala					7561
_	_	_		_	Cys	gga Gly	_	_		Val	_		_	_	Ser	7609
				Glu		gag Glu			Leu					Glu		7657
			${\tt Tyr}$			ccc Pro		Gly.					Pro			7705
		Glu				tca Ser 1975	Cys					Ser				7753
	Ala					gtg Val )					Arg					7801
ccc Pro	ctt Leu	gcg Ala	cgg Arg	gct Ala 2009	Ala	tgg Trp	gag Glu	aca Thr	gct Ala 2010	Arg	cac His	act Thr	cca Pro	gtc Val 2019	Asn	7849
				Asn		atc Ile			Ala					Ala		7897
			Met			ttc Phe		Ser					Gln			7945
		Lys				tgt Cys 2055	Gln					Cys				7993
gag Glu 206	Pro	ctt Leu	gac Asp	cta Leu	cct Pro 2070	Gln	atc Ile	att Ile	caa Gln	cga Arg 2075	Leu	cac His	Gly ggc	ctt Leu	agc Ser 2080	8041
					Ser	tac Tyr				Glu					Ala	8089

2100 2105 2110	8137
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	8185
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	8233
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	8281
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	8329
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 2190	8377
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	8427
ggccaatagg ccatcctgtt tttttcccct tttttttt tttttttc tttttttt	8547
3-3-3-3-03 addees coordenate cause	8642
<210> 3 <211> 2201 <212> PRT <213> HCV	
<210> 3 <211> 2201 <212> PRT	
<210> 3 <211> 2201 <212> PRT <213> HCV  <220> <221> VARIANT <222> 882	
<210> 3 <211> 2201 <212> PRT <213> HCV  <220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg  <221> VARIANT <222> 1489	
<pre>&lt;210&gt; 3 &lt;211&gt; 2201 &lt;212&gt; PRT &lt;213&gt; HCV  &lt;220&gt; &lt;221&gt; VARIANT &lt;222&gt; 882 &lt;223&gt; Xaa is Lys or Arg  &lt;221&gt; VARIANT &lt;222&gt; 1489 &lt;223&gt; Xaa is Leu  &lt;400&gt; 3 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1</pre>	
<pre> &lt;210&gt; 3 &lt;211&gt; 2201 &lt;212&gt; PRT &lt;213&gt; HCV  &lt;220&gt; &lt;221&gt; VARIANT &lt;222&gt; 882 &lt;223&gt; Xaa is Lys or Arg  &lt;221&gt; VARIANT &lt;222&gt; 1489 &lt;223&gt; Xaa is Leu  &lt;400&gt; 3  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1</pre>	
<pre> &lt;210&gt; 3</pre>	

#### 23 / 93

Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly 90 Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala 105 Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala 120 125 Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu 135 Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val 155 150 Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr 165 170 Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro 185 Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser 200 Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser 215 220 Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly 230 235 Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala 250 245 Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val 260 265 270 Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile 280 Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala 295 300 Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp 310 315 Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg 325 330 Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu 340 345 Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val 360 365 355 Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val 395 390 Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val 405 410 Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro 420 425 Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser 435 440 Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly 455 460 Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala 470 475 Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys 485 490 495 Ser Gly Gly Ala Tyr Asp Ile Ile Cys Asp Glu Cys His Ser Thr 505 Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu 520 525 Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly 535

#### 24 / 93

Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile 565 570 Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp 580 585 590 Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr 600 Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile 610 615 620 Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp 630 635 Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser 650 645 Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala 660 665 Val Ser Arg Ser Gln Arg Gly Arg Thr Gly Arg Gly Arg Met Gly 680 Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp 695 700 Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu 710 715 Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr 725 730 Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val 740 745 Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys 755 760 765 Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val 775 780 Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys 790 795 Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu 805 810 Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr His Pro Ile 825 820 Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr 840 Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr 855 860 Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser 870 875 Gly Xaa Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe 885 890 Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly 900 905 910 Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln 920 Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys 940 935 Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile 950 955 Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro 965 970 Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser. Pro Leu 980 985 990 Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala 995 1000 1005

#### 25 / 93

Ala Gln Leu Ala Pro Pro Ser Ala Ala Ser Ala Phe Val Gly Ala Gly 1015 1020 Ile Ala Gly Ala Ala Val Gly Ser Ile Gly Leu Gly Lys Val Leu Val 1030 1035 . Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala 1045 1050 1055 Phe Lys Val Met Ser Gly Glu Met Pro Ser Thr Glu Asp Leu Val Asn 1060 · 1065 1070 Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val 1075 1080 1085 Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val 1090 1095 1100 Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val 1105 1110 1115 Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr 1125 1130 1135 Gln Ile Leu Ser Ser Leu Thr Ile Thr Gln Leu Leu Lys Arg Leu His 1145 1150 1140 Gln Trp Ile Asn Glu Asp Cys Ser Thr Pro Cys Ser Gly Ser Trp Leu 1155 1160 1165 Arg Asp Val Trp Asp Trp Ile Cys Thr Val Leu Thr Asp Phe Lys Thr 1170 1175 1180 Trp Leu Gln Ser Lys Leu Leu Pro Arg Leu Pro Gly Val Pro Phe Phe 1190 1195 1200 Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met 1205 1210 Gln Thr Thr Cys Pro Cys Gly Ala Gln Ile Thr Gly His Val Lys Asn 1220 1225 1230 Cys Ser Met Arg Ile Val Gly Pro Arg Thr Cys Ser Asn Thr Trp His 1235 1240 1245 Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Ser 1250 1255 1260 Pro Ala Pro Asn Tyr Ser Arg Ala Leu Trp Arg Val Ala Ala Glu Glu 1265 1270 1275 1280 Tyr Val Glu Val Thr Arg Val Gly Asp Phe His Tyr Val Thr Gly Met 1285 1290 1295 Thr Thr Asp Asn Val Lys Cys Pro Cys Gln Val Pro Ala Pro Glu Phe 1300 1310 1305 Phe Thr Glu Val Asp Gly Val Arg Leu His Arg Tyr Ala Pro Ala Cys 1315 1320 1325 Lys Pro Leu Leu Arg Glu Glu Val Thr Phe Leu Val Gly Leu Asn Gln 1330 1335 1340 Tyr Leu Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala 1345 1350 1355 Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr 1365 1370 1375 Ala Lys Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Leu Ala Ser Ser 1380 1385 . 1390 Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Thr 1395 1400 1405 Arg His Asp Ser Pro Asp Ala Asp Leu Ile Glu Ala Asn Leu Leu Trp 1415 1420 Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys 1430 1435 Val Val Ile Leu Asp Ser Phe Glu Pro Leu Gln Ala Glu Glu Asp Glu 1445 1450 Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Arg Ser Arg Lys Phe 1460 1465 1470

#### 26 / 93

Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu 1480 1475 Xaa Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly 1490 1495 1500 Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg 1505 1510 1515 **1**520 Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala 1525 1530 Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp 1540 1545 1550 Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp 1555 1560 1565 Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly 1570 1575 1580 Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser 1590 1595 1600 Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp 1610 1615 1605 Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro 1620 1625 1630 Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr 1635 1640 1645 Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe 1655 1660 Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu 1675 1670 Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu 1685 1690 1695 Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly 1700 1705 1710 Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His 1715 1720 1725 Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile 1730 1735 1740 Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu 1750 1755 1760 1745 Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly 1765 1770 Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu 1780 1785 1790 Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly 1800 1795 1805 Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro 1815 1820 Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu 1825 1830 1835 1840 Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala 1845 1850 Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile 1865 1860 1870 Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg 1875 1880 1885 Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr 1895 1900 Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp 1910 1915 1920 Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser 1925 1930

```
Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala
                                                 1950
                              1945
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr
                                             1965
       1955
                         1960
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His
   1970
                     1975
                                        1980
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr
                 1990
                                     1995
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn
              2005 2010 2015
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg
          2020
                              2025
                                                2030
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln
                          2040
                                            2045
       2035
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile
                      2055
                                         2060
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser
                  2070
                                     2075
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala
              2085
                                 2090
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His
                           2105
                                                2110
          2100
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala
                         2120
                                            2125
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
   2130
                      2135
                                     2140
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
                  2150
2145
                                    2155
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
               2165
                                 2170
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly
           2180
                             2185
Val Gly Ile Tyr Leu Leu Pro Asn Arg
       2195
                          2200
<210> 4
<211> 8643
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
accagecece gattggggge gacaétecae catagateae teceetgtga ggaaetaetg 60
tetteaegea gaaagegtet ageeatggeg ttagtatgag tgtegtgeag ceteeaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gateaacceg eteaatgeet ggagatttgg gegtgeecee 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
eggeegettg ggtggagagg etattegget atgaetggge geaacagaea ateggetget 480
etgatgeege egtgtteegg etgteagege aggggegee ggttettttt gteaagaeeg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tecttgcgca getgtgetcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
```

ttgff ccag gctf tggg ttgg agcg ccgg gccg tagg agtf gaad tgca atgg tatg	cegat ggete ggete gegge geate ggete teece aaag getet gggat teece aaag getet gggat teeg teg ggat teeg gggat gggat gggat gggat teece t	cca graa figa figa figa figa figa figa figa fi	ggatgggggggggggggggggggggggggggggggggg	gatetegeatagetegeatagetegeatagetegeatagetegeatagetegeatagetegeatagetegeatagetegeatagegeatagegeatagegeatagegeatagegeatagegeatagegeatagegeatagegeatageatagegeatagea	eg ga eg co eg ga at ca eg ga at ca eg ga eg co est ga eg co est ga eg co est ga eg co est ga eg co est ga eg co est ga est co est ga est co est ga est co est ga est co est ga est g est g est ga est g est	acgaa ccgac aaaat aggac gcttc tctt agccggt ggccc caaa gccaggt gcca caac acggt ccaggt	agago cggco catago cgaco ggaat cgtgo cggaa ggaat catago cgcaco ggga cggaa	ato ago ago ago ago ago ago ago ago ago ag	caggggatch ctttc cttgg ctttc cttgt cctggg caggt cgagg cgaag	ggct ccgt ctgg ctac acgg cctg cctg cctg	cgcg cgtg atto ccgt tato agti cccc atgt gcga atag gcga tttg	gocaç gacaç gata cgata cgata cgccccccccc tatt tatcctt accct gaga accct accc	gee geat gat gat gat gat gat gat gat gat gat g	yaactygegacgacgataaaggacgataaaggacgataaaggacgacgataaaggacgataaggataaggacgataaggacgataaggacgataaggacgataaggacgataaggacgataaggacgataaggacgataaggacgataaggacgataaggacgataaggataaggataaggataaggataagg	eggte gtteg atgeet geegge attege attee gttaa actgg catatt geageg acacc agtaa gttaa gttaa gttaa gttaa gttaa gttaa g gtaa g gtaa gtaa g gt	900 960 1020 1080 1140 1200 1320 1320 1440 1560 1620 1680 1740 1800
_			ttg Leu 20		_		_			_	_			_		1897
			tgg Trp													1945
			atc Ile													1993
			acg Thr	_												2041
			ctc Leu													2089
			gtg Val 100													2137
		_	gtg Val	_	_					•	_		${\tt Gln}$			2185
			ttg Leu													2233
			cgg Arg													2281

												gtt Val			2329
												ggc Gly 190			2377
												gca Ala			2425
												gcc Ala			2473
	_	_	_					_			_	ctc Leu			2521
	_			_	_				_		_	tcc Ser		_	2569
												tgg Trp 270			2617
												ggc			2665
		_					_	_	_	-		tgg Trp			2713
				_		_			_	_		agc Ser	_	-	2761
				_								cgc Arg			2809
	_	_			_						-	tcc Ser 350		_	2857
_			-				_		_	_		cac His	_		2905
												aag Lys			2953

				gtc Val											3001
				tcg Ser 405											3049
_				gcc Ala			_		_	_		_		_	3097
				gcc Ala											3145
				cta Leu											3193
				atc Ile											3241
				tcc Ser 485											3289
			_	tat Tyr	-			_	_		-				3337
				atc Ile											3385
				cga Arg											3433
_	_			cca Pro							_	_		_	3481
				ccc Pro 565											3529
				cac His											3577
_		_		aag Lys	-						_	_	_		3625

	 		-	_		_			_		gac Asp	_		3673
											gat Asp			3721
											gac Asp			3769
			Phe								caa Gln 670			3817
	_	_	_		_						agg Arg	_		3865
											atg Met			3913
											tgg Trp			3961
											cta Leu			4009
											gag Glu 750			4057
						_	_		_		cag Gln		_	4105
											gct Ala			4153
											atg Met			4201
_				•		-	•			_	ccc Pro	-	_	4249
											cac His 830			4297

acc aaa tac atc Thr Lys Tyr Ile 835	atg gca tgc Met Ala Cys	atg tcg gct Met Ser Ala 840	gac ctg gag gtc Asp Leu Glu Val 845	gtc acg 4345 Val Thr
agc acc tgg gtg Ser Thr Trp Val 850	ctg gta ggc Leu Val Gly 855	gga gtc cta Gly Val Leu	gca gct ctg gcc Ala Ala Leu Ala 860	gcg tat 4393 Ala Tyr
			ggc agg atc atc Gly Arg Ile Ile 875	
			gtc ctt tac cgg Val Leu Tyr Arg	
	Glu Cys Ala		cct tac atc gaa Pro Tyr Ile Glu 910	
	_		gca atc ggg ttg Ala Ile Gly Leu 925	_
			ccc gtg gtg gaa Pro Val Val Glu 940	_
			cat atg tgg aat His Met Trp Asn 955	
			act ctg cct ggc Thr Leu Pro Gly	
			tct atc acc agc Ser Ile Thr Ser 990	
			ctg ggg gga tgg Leu Gly Gly Trp 1005	
		Ala Ala Ser	gct ttc gta ggc Ala Phe Val Gly 1020	
			ctt ggg aag gtg Leu Gly Lys Val 1035	
			gca ggc gcg ctc Ala Gly Ala Leu O	

ttt aag gtc atg Phe Lys Val Met 106	Ser Gly Glu	atg ccc tcc Met Pro Ser 1065	Thr Glu Asp Le	ng gtt aac 5017 eu Val Asn 070
cta ctc cct gct Leu Leu Pro Ala 1075				
tgc gca gcg ata Cys Ala Ala Ile 1090	ctg cgt cgg Leu Arg Arg 109	His Val Gly	cca ggg gag gg Pro Gly Glu G 1100	gg gct gtg 5113 ly Ala Val
cag tgg atg aac Gln Trp Met Asn 1105	cgg ctg ata Arg Leu Ile 1110	gcg ttc gct Ala Phe Ala	tcg cgg ggt a Ser Arg Gly A 1115	ac cac gtc 5161 sn His Val 1120
tcc ccc acg cac Ser Pro Thr His	tat gtg cct Tyr Val Pro 1125	gag agc gac Glu Ser Asp 113	Ala Ala Ala A	gt gtc act 5209 rg Val Thr 1135
cag atc ctc tct Gln Ile Leu Ser 114	Ser Leu Thr	_	Leu Leu Lys A	<del>-</del>
cag tgg atc aac Gln Trp Ile Asn 1155				
aga gat gtt tgg Arg Asp Val Trp 1170	gat tgg ata Asp Trp Ile 117	Cys Thr Val	ttg act gat to Leu Thr Asp Pl 1180	tc aag gcc 5353 he Lys Ala
tgg ctc cag tcc Trp Leu Gln Ser 1185				
tca tgt caa cgt Ser Cys Gln Arg			Arg Gly Asp G	
caa acc acc tgc Gln Thr Thr Cys 122	Pro Cys Gly		Thr Gly His V	
tgt tcc atg agg Cys Ser Met Arg 1235				
gga aca ttc ccc Gly Thr Phe Pro 1250		Tyr Thr Thr		
ccg gcg cca aat Pro Ala Pro Asn 1265	tat tct agg Tyr Ser Arg 1270	gcg ctg tgg Ala Leu Trp	cgg gtg gct g Arg Val Ala A 1275	ct gag gag 5641 la Glu Glu 1280

	gtg Val				Arg					His					Met	5689
	act Thr			Val					Gln					Glu		5737
	aca Thr		Val					Leu					Pro			5785
aaa Lys	ccc Pro 1330	Leu	cta Leu	cgg Arg	gag Glu	gag Glu 1339	Val	aca Thr	ttc Phe	ctg Leu	gtc Val 1340	Gly	ctc Leu	aat Asn	caa Gln	5833
	ctg Leu 5					Leu					Glu					5881
	ctc Leu				Leu					His					Thr	5929
	aag Lys	_		Leu	_				Pro			_	_	Ser		5977
	gct Ala		Gln					Ser					Cys			6025
	cat His 1410	Asp					Asp					Asn				6073
	cag Gln					Asn					Glu					6121
gta Val	gta Val	att Ile	ttg Leu	gac Asp 1445	Ser	ttc Phe	gag Glu	ccg Pro	ctc Leu 1450	Gln	gcg Ala	gag Glu	gag Glu	gat Asp 1455	Glu	6169
	gaa Glu			Val					Leu					Lys		6217
	cga Arg		Met					Arg					Pro			6265
	gag Glu 1490	Ser		_		_	Asp		_			Val	-			6313

tgt Cys 150!	Pro	ttg Leu	ccg Pro	cct Pro	gcc Ala 151	ГЛЗ	gcc Ala	cct Pro	ccg Pro	ata Ile 151		cct Pro	cca Pro	cgg Arg	agg Arg 1520	6361
			gtt Val		Leu					Val					Ala	6409
			aca Thr 1540	Lys					Ser					Val		6457
			gca Ala 5					Asp					Asp			6505
gcg Ala	gga Gly 1570	Ser	gac Asp	gtt Val	gag Glu	tcg Ser 157	Tyr	tcc Ser	tcc Ser	atg Met	ccc Pro 158	Pro	ctt Leu	gag Glu	<b>GJ</b> Å aaa	6553
gag Glu 1589	Pro	gjå aaa	gat Asp	ccc Pro	gat Asp 1590	Leu	agc Ser	gac Asp	gly aaa	tct Ser 159	$\operatorname{Trp}$	tct Ser	aċc Thr	gta Val	agc Ser 1600	6601
			agt Ser		qeA					Ser					Trp	6649
acg Thr	Gly	gcc Ala	ctg Leu 1620	Ile	acg Thr	cca Pro	tgc Cys	gct Ala 1625	Ala	gag Glu	gaa Glu	acc Thr	aag Lys 1630	Leu	ccc Pro	6697
			ctg Leu 5					Leu					Leu			6745
gct Ala	aca Thr 1650	Thr	tct Ser	cgc Arg	agc Ser	gca Ala 1655	Ser	ctg Leu	cgg Arg	cag Gln	aag Lys 1660	Lys	gtc Val	acc Thr	ttt Phe	6793
gac Asp 1665	Arg	ctg Leu	cag Gln	gtc Val	ctg Leu 1670	Asp	gac Asp	cac His	Tyr	cgg Arg 1675	Asp	gtg Val	ctc Leu	aag Lys	gag Glu 1680	6841
			aag Lys		Ser					Lys					Glu	6889
gaa Glu	gcc Ala	tgt Cys	aag Lys 1700	Leu	acg Thr	ccc Pro	cca Pro	cat His 1705	Ser	gcc Ala	aga Arg	tct Ser	aaa Lys 1710	Phe	Gl <sup>à</sup> aac	6937
			aag Lys					Leu					Val			6985

atc of I		Ser					Leu					Glu				7033
gac a Asp :						Lys					Cys					7081
aag g Lys (			-	_	Pro	_	-			Val			_	_	Gly	7129
gtt o	_		_	Glu		-			Tyr	_		_		Thr		7177
cct o			Val					$\mathtt{Tyr}$					Ser			7225
cag o		Val					Asn					Lys				7273
atg 6 Met 0 1825						Thr					Ser					7321
aat g Asn <i>I</i>					Glu					Gln					Ala	7369
ccc g Pro (	-		-	Gln	_			_	Leu					Tyr		7417
Gly o			Leu													7465
							1880	_		Asn	Cys	1885	_		J	
tgc ( Cys I		Ala	agc				acg Thr	acc	agc	tgc	ggt	1889 aat Asn	acc	ctc	aca	7513
Cys I	Arg 1890 tac	Ala	agc Ser	Gly	Val gct	Leu 1895 gcg Ala	acg Thr	acc Thr	agc Ser cga	tgc Cys gct	ggt Gly 1900 gcg Ala	aat Asn	acc Thr	ctc Leu cag	aca Thr	7513 7561
Cys I tgt t	Arg 1890 tac Tyr	Ala ttg Leu atg	agc Ser aag Lys	Gly gcc Ala gta	yal gct Ala 1910 tgc Cys	Leu 1895 gcg Ala )	acg Thr gcc Ala	acc Thr tgt Cys	agc Ser cga Arg	tgc Cys gct Ala 1915 gtc Val	ggt Gly 1900 gcg Ala 5	aat Asn aag Lys	acc Thr ctc Leu	ctc Leu cag Gln	aca Thr gac Asp 1920 agc Ser	

Mec	act Thr		Tyr					Gly					Pro			7705
_	ttg Leu 1970	Glu	_				Cys					Ser	_			7753
	gca Ala 5					Val					Arg					7801
	ctt Leu	-		_	Ala				-	Arg				_	Asn	7849
	tgg Trp			Asn					Ala					Ala		7897
_	atc Ile	_	Met					Ser					Gln	-		7945
	gaa Glu 2050	Lys					Gln					Cys				7993
gag	cca	ctt	~~~	cta	cct		240	n++			a+ a		~~~	a++		8041
Glu 206!	Pro		_			Gln				_	Leu				_	9041
2069 gca	Pro	Leu tca	Asp	Leu cat	Pro 2070 agt Ser	Gln ) tac	Ile tct	Ile	Gln ggt.	Arg 2079 gag Glu	Leu S atc	His aat	Gly	Leu gtg	Ser 2080 gct Ala	8089
gca Ala tca	Pro ttt	Leu tca Ser	Asp ctc Leu	cat His 2089 aaa Lys	Pro 2070 agt Ser ctt	Gln tac Tyr	Ile tct Ser	cca Pro	ggt Gly 2090 ccc Pro	Arg 2079 gag Glu ttg	Leu atc Ile cga	His aat Asn gtc	Gly agg Arg	gtg Val 2099 aga Arg	gct Ala	
gca Ala tca Ser	Pro ttt Phe tgc	tca Ser ctc Leu	Asp ctc Leu agg Arg 2100 agt Ser	cat His 2089 aaa Lys	Pro 2070 agt Ser ctt Leu	Gln tac Tyr ggg Gly gct	tct Ser gta Val	cca Pro ccg Pro 2109	ggt Gly 2090 ccc Pro	Arg 2075 gag Glu ttg Leu	atc Ile cga Arg	His aat Asn gtc Val	Gly agg Arg tgg Trp 2110 ggg Gly	gtg Val 2099 aga Arg	gct Ala cat His	8089
gca Ala tca ser cgg Arg	Pro ttt Phe tgc Cys	tca Ser ctc Leu aga Arg 2115 tgt	Asp ctc Leu agg Arg 2100 agt Ser	cat His 2089 aaa Lys gtc Val	Pro 2070 agt Ser ctt Leu cgc Arg	tac Tyr ggg Gly gct Ala	tct Ser gta Val agg Arg 2120 ttc Phe	cca Pro ccg Pro 2109 cta Leu	ggt Gly 2090 ccc Pro ctg Leu	Arg 2079 gag Glu ttg Leu tcc Ser	atc Ile cga Arg cag Gln	aat Asn gtc Val ggg Gly 2125 agg Arg	agg Arg tgg Trp 2110 ggg Gly	gtg Val 2099 aga Arg agg Arg	gct Ala cat His gct Ala	8089
gca Ala tca ser cgg Arg gcc Ala	ttt Phe tgc Cys gcc Ala act Thr 2130	tca ser ctc Leu aga Arg 2115 tgt Cys	Asp ctc Leu agg Arg 2100 agt ser ggc Gly	cat His 2089 aaa Lys gtc Val aag Lys	ero 2070 agt ser ctt Leu cgc Arg	tac Tyr ggg Gly gct Ala ctc Leu 2135	tct ser gta Val agg Arg 2120 ttc Phe	cca Pro ccg Pro 2109 cta Leu aac Asn	ggt Gly 2090 ccc Pro ctg Leu tgg	gag Glu ttg Leu tcc ser gca Ala	atc Ile cga Arg cag Gln gta Val 2140 gat Asp	aat Asn gtc Val ggg Gly 2125 agg Arg	agg Arg tgg Trp 2110 ggg Gly acc Thr	gtg Val 2099 aga Arg agg Arg aag	gct Ala cat His gct Ala ctc Leu	8089 8137 8185

#### 38 / 93

gee ega eee ege tgg tte atg tgg tge eta ete eta ett tet gta ggg 8377 Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 2190 gta ggc atc tat cta ctc ccc aac cga tga acggggaget aaacactcca 8427 Val Gly Ile Tyr Leu Leu Pro Asn Arg \* 2195 2200 titttttttt tttttttt ttttctttt tcccaatttt tttccttttc tttcctttgg 8547 tggctccatc ttagccctag tcacggctag ctgtgaaagg tccgtgagcc gcttgactgc 8607 agagagtgct gatactggcc tctctgcaga tcaagt <210> 5 <211> 8648 <212> DNA <213> HCV <220> <221> CDS <222> (1802)...(8407) <400> 5 gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60 tetteaegea gaaagegtet ageeatggeg ttagtatgag tgtegtgeag eetceaggae 120 ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180 gacgaceggg teetttettg gateaacecg etcaatgeet ggagatttgg gegtgeeece 240 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300 gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360 ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420 cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480 etgatgeege egtgtteegg etgteagege aggggegeee ggttettttt gteaagaeeg 540 acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600 cgacgggcgt teettgegea getgtgeteg aegttgteae tgaageggga agggaetgge 660 tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720 aagtateeat catggetgat geaatgegge ggetgeatae gettgateeg getacetgee 780 cattcgacca ccaagcgaaa catcgcatcg agcgagcacg tactcggatg gaagccggtc 840 ttqtcqatca ggatgatctq gacqaaqaqc atcaqqqqct cqcqccaqcc gaactqttcq 900 ccaggeteaa ggegegeatg cccgaeggeg aggatetegt egtgaeceat ggegatgeet 960 gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020 tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080 ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140 agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200 cagaccacaa eggttteect etagegggat caatteegee eeeeceeta aegttaetgg 1260 ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320 gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380 taggggtett teceeteteg ceaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440 agtteetetg gaagettett gaagacaaac aaegtetgta gegaceettt geaggeageg 1500 gaacccccca cetggegaca ggtgeetetg eggecaaaag ecaegtgtat aagatacaec 1560 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620 atggetetee teaagegtat teaacaaggg getgaaggat geecagaagg taccecattg 1680 tatgggatet gatetgggge eteggtgeae atgetttaea tgtgtttagt egaggttaaa 1740 aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800 c atg gac egg gag atg gca gca teg tgc gga ggc geg gtt ttc gta ggt 1849 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly

										•						
	_			_		_		_	cac His 25		_	_			-	 1897
									atc Ile							1945
									gtt Val							1993
				_	_				cca Pro				•			2041
			_		_				cca Pro		_			_	_	 2089
									cgc Arg 105							2137
									gjå aaa							2185
		_	_	_	_	_	_		ggt Gly	_		_		_		2233
•			_		-		_		gcg Ala			_	-			 2281
	_	_			_	_			gat Asp	_			_	_		2329
			_	_				-	999 Gly 985	_			_		_	2377
							Arg		ata Ile							2425
									ctc Leu							2473
1									tgc Cys							2521

										gtg Val				2569
										gtg Val				2617
			_		_			_		cca Pro	_			2665
										gtc Val 300				2713
										tgc Cys				2761
		_	_	_		_	_	_		ccg Pro		_		 2809
										ccc Pro				2857
_			_			_		_		tcg Ser			_	 2905
										gtt Val 380				2953
										atg Met				3001
	-	_		-		-	-	_	_	cag Gln			_	 3049
										agc Ser				3097
										gtc Val				3145
										tct Ser 460				3193

	gac Asp															3241
	atc Ile															3289
	gjå aaa															3337
	tcg Ser															3385
	gct Ala 530															3433
	gtc Val															3481
	gga Gly	_							_							3529
_	gly aaa							_			_	_		_	_	3577
	ctc Leu	-		_	_							_	_	_		3625
	cgg Arg 610															3673
-	gta Val	-	-	_	_		_	_					_		_	3721
	gtg Val															3769
ctg Leu	gac Asp	ccg Pro	acc Thr 660	ttc Phe	acc	att Ile	gag Glu	acg Thr 665	acg Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	gac Asp	gcg Ala	3817
gtg Val	tca Ser	cgc Arg 675	tcg Ser	cag Gln	cgg Arg	cga Arg	Gly Ggc	agg Arg	act Thr	ggt Gly	agg Arg	ggc Gly 685	agg Arg	atg Met	ggc ggc	3865

							42 /	93						
			ttt Phe				gaa	cgg						3913
			ctg Leu											3961
	~		gcc Ala			_		_		_				4009
			ccc Pro 740											4057
			ctc Leu											4105
			gac Asp											4153
			gct Ala											4201
_			cgg Arg		_	_	_				_		_	4249
		-	gga Gly 820	_	_			_						4297
			atc Ile											4345
		Trp	gtg Val		Val	Gly			Ala					4393
			aca Thr											4441
			gcc Ala											4489
			gaa Glu 900											4537

												ggg Gly 925				4585
												gtg Val				4633
tgg Trp 945	cgg Arg	acc Thr	ctc Leu	gaa Glu	gcc Ala 950	ttc Phe	tgg Trp	gcg Ala	aag Lys	cat His 955	atg Met	tgg Trp	aat Asn	ttc Phe	atc Ile 960	4681
												cct Pro				4729
				Leu								acc Thr				4777
			His					Asn				gga Gly 1005	Trp			4825
		Leu					Ala					gta Val )				4873
	Ala					Gly					Gly	aag Lys				4921
Ile 1029 gat	Ala att	Gly ttg	Ala	Ala ggt	Val 1030 tat Tyr	Gly gga	Ser	Ile	gtg	Leu 1039 gca Ala	ggc Gly		'Val ctc	Leu gtg	Val 1040 gcc Ala	4921 4969
Ile 1029 gat Asp	Ala att Ile	Gly ttg Leu gtc	Ala gca Ala	Ala ggt Gly 104! agc Ser	Val 1030 tat Tyr 5	Gly gga Gly gag	gca Ala atg	ecc aaa Ile	gtg Val 1050 tcc ser	Leu 1039 gca Ala )	Gly ggc Gly	gcg	val ctc Leu ctg	gtg Val 1059 gtt Val	Val 1040 gcc Ala	
Ile 1025 gat Asp ttt Phe	att Ile aag Lys ctc Leu	ttg Leu gtc Val	gca Ala atg Met 1060 gct Ala	Ala ggt Gly 104! agc Ser	Val 1030 tat Tyr ggc Gly	gga Gly gag Glu	gca Ala atg Met	ggg Gly	gtg Val 1050 tcc ser	gca Ala acc Thr	ggc Gly gag Glu gtc	Lys gcg Ala gac	ctc Leu ctg Leu 1070 ggg Gly	gtg Val 1055 gtt Val )	yal 1040 gcc Ala aac Asn	4969
gat Asp ttt Phe cta Leu	Ala att Ile aag Lys ctc Leu gca	ttg Leu gtc Val cct Pro 107! gcg Ala	gca Ala . atg Met 106 gct Ala 5	Ala ggt Gly 104! agc ser o atc Ile	Val 1030 tat Tyr ggc Gly ctc Leu	gga Gly gag Glu tcc ser	gca Ala atg Met cct Pro 1080	ggc Gly ggc Gly	gtg Val 1050 tcc Ser gcc Ala	gca Ala acc Thr	ggc gly gag Glu gtc Val	gcg Ala gac Asp gtc Val 1089 gag Glu	ctc Leu ctg Leu 1070 ggg Gly 5	gtg Val 1055 gtt Val gtc Val	Val 1040 gcc Ala aac Asn gtg Val	4969 5017
gat Asp ttt Phe cta Leu tgc Cys	att Ile aag Lys ctc Leu gca Ala 1090 tgg	gtc Val cct Pro 107! gcg Ala	gca Ala atg Met 106 gct Ala ata Ile	Ala . ggt Gly 104! agc Ser . atc Ile ctg Leu cgg	tat Tyr ggc Gly ctc Leu cgt Arg	gga Gly gag Glu tcc ser cgg Arg 1099	gca Ala atg Met cct Pro 1080 cac His	ggg Gly ccc Pro 1069 ggc Gly gtg Val	gtg Val 1050 tcc ser 5 gcc Ala	gca Ala acc Thr cta Leu cca Pro	ggc gly gag gly 1100 cgg Arg	gcg Ala gac Asp gtc Val 1089 gag Glu	ctc Leu ctg Leu 1070 ggg Gly 6	gtg Val 1059 gtt Val gtc Val gtc Ala	Val 1040 gcc Ala aac Asn gtg Val gtg Val	4969 5017 5065

			cag ctg ctg Gln Leu Leu 5		
	Asn Glu Asp		cca tgc tcc Pro Cys Ser		
			gtg ttg act Val Leu Thr 118	Asp Phe Lys	
		Leu Pro Arg	ttg ccg gga Leu Pro Gly 1195		
			tgg cgg ggc Trp Arg Gly 1210		Met
			atc acc gga Ile Thr Gly 5		
_	Arg Ile Val		acc tgt agt Thr Cys Ser		
			acg ggc ccc Thr Gly Pro 1260	Cys Thr Pro	
		Arg Ala Leu	tgg cgg gtg Trp Arg Val 1275		
			ttc cac tac Phe His Tyr 1290		Met
_			cag gtt ccg Gln Val Pro 5	-	
	Val Asp Gly		cac agg tac His Arg Tyr		
			ttc ctg gtc Phe Leu Val . 1340	Gly Leu Asn	
		Leu Pro Cys	gag ccc gaa Glu Pro Glu 1355		

gtg o				_	Leu		_			His		-		_	Thr	5929
gct a Ala I				Leu					Pro					Ser		5977
tca g Ser A			Gln					Ser					Cys			6025
cgt o		Asp		_	_	_	Asp				_	Asn		_		6073
cgg o Arg 0 1425	Gln		_			Asn			_		Glu		_		_	6121
gta g Val V					Ser					Gln					Glu	6169
agg g Arg (				Val					Leu					Lys		6217
ect o			Met					Arg					Pro			6265
tta g Leu G		Ser		_	_	-	Āsp		-			Val	-			6313
tgt o Cys I 1505	Pro					Lys					Pro					6361
aag a Lys A		_	_	_	Leu		_			Val			~	_	Ala	6409
gag d Glu I				Lys					Ser					Val		6457
agc g Ser (		_	Ala		-			Asp				_	Asp		_	6505
gcg g Ala G		Ser					Tyr					Pro				6553

gag ccg ggg Glu Pro Gly 1585		Leu Ser			_	-
gag gag gct Glu Glu Ala			_	Ser Met Se		Trp
aca ggc gcc Thr Gly Ala		Pro Cys				
atc aat gca Ile Asn Ala 1635	Leu Ser Asn		Leu Arg	His His As		
gct aca aca Ala Thr Thr 1650						
gac aga ctg Asp Arg Leu 1665		Asp Asp				
atg aag gcg Met Lys Ala				Lys Leu Le		Glu
gaa gcc tgt Glu Ala Cys		Pro Pro				
tat ggg gca Tyr Gly Ala 1715	Lys Asp Val		Leu Ser	Ser Lys Al		
atc cgc tcc Ile Arg Ser 1730	gtg tgg aag Val Trp Lys	gac ttg Asp Leu 1 1735	ctg gaa Leu Glu	gac act ga Asp Thr Gl 1740	g aca cca u Thr Pro	att 7033 Ile
gac acc acc Asp Thr Thr 1745		Lys Asn	Glu Val			
aag ggg ggc Lys Gly Gly				Val Phe Pr		Gly
gtt cgt gtg Val Arg Val	tgc gag aaa Cys Glu Lys 1780	Met Ala	ctt tac Leu Tyr 1785	gat gtg gt Asp Val Va	c tcc acc l Ser Thr 1790	ctc 7177 Leu
cct cag gcc Pro Gln Ala 1795	Val Met Gly		Tyr Gly	Phe Gln Ty		

													-			
		Val					Asn				gcg Ala 1820	Lys				7273
	Gly					Thr					tca Ser					7321
					Glu					Gln	tgt C <b>y</b> s				Ala	7369
				Gln					Leu		gag Glu			Tyr		7417
			Leu					Gly			tgc Cys		Tyr			7465
		Ala					Thr				ggt Gly 1900	Asn				7513
	Tyr					Ala					gcg Ala 5					7561
_	_	_		_	Cys		_	_		Val	gtt Val		_	_	Ser	7609
				Glu					Leu		gcc Ala			Glu		7657
			Tyr					Gly			ccc Pro		Pro			7705
		Glu					Cys				gtg Val 1980	Ser				7753
	Ala					Val					cgt Arg					7801
					Ala					Arg	cac His				Asn	7849
tcc Ser	tgg Trp	cta Leu	ggc Gly 2020	Asn	atc Ile	atc Ile	atg Met	tat Tyr 2029	Ala	ccc Pro	acc Thr	ttg Leu	tgg Trp 2030	Ala	agg Arg	7897

# 48 / 93

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 79 Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln 2035 2040 2045	945
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile 2050 2055 2060	993
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 80 Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2065 2070 2075 . 2080	041
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 80 Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 2095	089
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 85 Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	137
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8: Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	185
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 82 Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	233
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 82 Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	281
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 83 Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	329
gcc cga ccc cgc tgg ttc acg tgg tgc cta ctc cta ctt tct gta ggg 83 Ala Arg Pro Arg Trp Phe Thr Trp Cys Leu Leu Leu Ser Val Gly 2180 · 2185 2190	377
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 84 Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	427
ggccaatagg ccatcetgtt tttttccctt ttttcccttt tttttttt tttttt	547
<210> 6 <211> 8638 <212> DNA <213> HCV <220> <221> CDS <222> (1802)(8407)	

<400> 6

	_	-	-		_				_			-	_		ctactg	
															caggac	
cccc	caato	cc g	gggag	gagco	ca ta	agtg	gtete	g cgg	gaac	cggt	gagt	caca	ccg (	gaati	tgccag	180
															gccccc	
gcga	agact	gc t	tagco	gagt	a gi	tgtt	gggto	geg	gaaag	ggcc	ttgi	ggta	act 9	gcct	gatagg	,300
gtg	cttgc	ga g	gtgco	ccgg	gg ag	ggtci	tcgta	a gad	ccgt	gcac	cate	gagca	acg a	aatc	ctaaac	360
ctca	aaaga	iaa a	aacca	aagg	gg cg	gege	catga	a ttg	gaaca	aaga	tgga	attg	cac g	gcag	gttctc	420
cggd	ceget	tg g	ggtgg	gagag	g ct	tatte	egget	: atg	gact	gggc	acaa	acaga	aca a	atcg	gctgct	480
ctga	atgco	gc (	cgtgt	tece	g ct	tgtc	agcgo	agg	gggcg	gccc	ggtt	ctti	tt ,	gtcaa	agaccg	540
acct	gtco	gg t	tgccc	ctgaa	at ga	aact	gcago	g ac	gagg	cagc	geg	gctai	cg '	tggc	tggcca	600
cgad	gggc	gt 1	tcctt	gege	a g	ctgt	gctc	g ac	gttgi	tcac	tgaa	agcg	gga	aggga	actggc	660
_		-			_			-			_				ccgaga	
_	_											-	_		cctgcc	
_				_					-		_	-	_	_	ccggtc	
	_		_	_		_	_	_		_			_	_	tgttcg	
															atgcct	
															geegge	
_								-				-	-		aagagc	
															attcgc	
		_		-	-	_	_	-				-	-	_	tgttaa	
															tactgg	
															catatt	
															cattcc	
	_	-			-										ggaagc	
															gcagcg	
aart																
						g cg c										
gaad						*+~~	7277									
gaad tgca	aaagg	cg g	gcaca	acco	cc ag	_	_	_			-	-		_	_	
gaad tgca atgg	aaagg gctct	cc t	gcaca tcaaç	acco	cc ag	caac	aaggg	g gct	tgaag	ggat	gcc	caga	agg	tacco	ccattg	1680
gaad tgca atgg tatg	aaagg gctct gggat	cc i	gcaca tcaag gatct	acco gcgta	ec ag at to gc ci	caac	aaggg tgċac	g gci	tgaag gctti	ggat taca	gcc	caga:	agg agt	tacco cgago	ccattg gttaaa	1680 1740
gaad tgca atgg tatg	aaagg getet gggat egtet	cg g cc t ct g	gcaca tcaag gatct gcccc	acco gegta geggg eccga	ec ag at to gc ci aa co	caaca ceggi	aaggg tgċad gggad	g gct c ato	tgaag getti ggtti	ggat taca ttcc	gccc tgtg ttttg	caga: yttt: yaaa:	agg agt aac	tacco cgagg acgal	ccattg gttaaa taatac	1680 1740 1800
gaad tgca atgg tatg aaad c at	aaagg getet gggat egtet eg ga	cg g cc g ct g ag g	gcaca tcaag gatct gccco gg ga	acco gegta geggg ecega ag at	ec agat to	caaca ceggi cacgg	aaggg tgċac gggac ca tc	g get e ato e gto eg to	tgaag getti ggtti ge gg	ggat taca ttcc ga gg	gcco tgtg tttg gc gc	caga gttt: gaaa cg g	agg agt aac act t	tacco cgago acgai tc gi	ccattg gttaaa taatac ta ggt	1680 1740 1800
gaad tgca atgg tatg aaad c at	aaagg getet gggat egtet eg ga et As	cg g cc g ct g ag g	gcaca tcaag gatct gccco gg ga	acco gegta ecega g at tu Me	ec agat to go of a co	caaca ceggi cacgg	aaggg tgċac gggac ca tc	g get e ato e gto eg to	tgaag getti ggtti ge gg ys G	ggat taca ttcc ga gg ly G:	gcco tgtg tttg gc gc	caga gttt: gaaa cg g	agg agt aac act t	tacco cgago acgai tc gi	ccattg gttaaa taatac ta ggt al Gly	1680 1740 1800
gaad tgca atgg tatg aaad c at	aaagg getet gggat egtet eg ga	cg g cc g ct g ag g	gcaca tcaag gatct gccco gg ga	acco gegta geggg ecega ag at	ec agat to go of a co	caaca ceggi cacgg	aaggg tgċac gggac ca tc	g get e ato e gto eg to	tgaag getti ggtti ge gg ys G	ggat taca ttcc ga gg	gcco tgtg tttg gc gc	caga gttt: gaaa cg g	agg agt aac act t	tacco cgago acgai tc gi	ccattg gttaaa taatac ta ggt	1680 1740 1800
gaad tgca atgg tatg aaad c at	aaagg getet gggat egtet eg ga et As	cc icc consider the consideration the considerat	geaca teaag gatet geece gg ga rg G]	acco gegta geggg eccga ag at	ec agat to ge et ac est A.	caaca ceggi cacgg ca go la A	aaggg tgċac gggac ca to la Se	g god c ato c gto eg to er C	tgaag getti ggtti ge gg ys G	ggat taca ttcc ga gg ly G:	gccc tgtg tttg gc gc	cagaa gttta gaaaa cg g la Va	agg agt aac at t	tacco cgagg acgat tc gt he Va	ccattg gttaaa taatac ta ggt al Gly 15	1680 1740 1800 1849
gaad tgca atgg tatg aaad c at Me	aaagg gctct gggat ggtct cg ga et As l ata	ctc	geace teaag gatet geece gg ge rg G]	acco gcgta gggg ccga ag at lu Me	ec agat to ge chan con ge ge ge t Al	caaca teggi caegg ca go la A	aaggg tgèac gggac ca to la Se	g got c atg c gtg cg tg er Cg	tgaaggetti ggtti gg gg gc gg ys G. :	ggat taca ttcc ga gg ly G: 10	gccc tgtg tttg gc gc ly A	cagaa gttta gaaaa cg gi la Va	agg agt aac at t al P	tacco cgagg acgat tc gt he Va	ccattg yttaaa taatac ta ggt al Gly 15	1680 1740 1800
gaad tgca atgg tatg aaad c at Me	aaagg getet gggat egtet eg ga et As	ctc	gcaca tcaac gatct gcccc gg ga rg G] ttg Leu	acco gcgta gggg ccga ag at lu Me	ec agat to ge chan con ge ge ge t Al	caaca teggi caegg ca go la A	aaggg tgèac gggac ca to la Se	g god c ato c gto cg to er Co cac His	tgaaggetti ggtti gg gg gc gg ys G. :	ggat taca ttcc ga gg ly G: 10	gccc tgtg tttg gc gc ly A	cagaa gttta gaaaa cg gi la Va	agg agt act t al P ctc Leu	tacco cgagg acgat tc gt he Va gct Ala	ccattg yttaaa taatac ta ggt al Gly 15	1680 1740 1800 1849
gaad tgca atgg tatg aaad c at Me	aaagg gctct gggat ggtct cg ga et As l ata	ctc	geace teaag gatet geece gg ga rg G]	acco gcgta gggg ccga ag at lu Me	ec agat to ge chan con ge ge ge t Al	caaca teggi caegg ca go la A	aaggg tgèac gggac ca to la Se	g got c atg c gtg cg tg er Cg	tgaaggetti ggtti gg gg gc gg ys G. :	ggat taca ttcc ga gg ly G: 10	gccc tgtg tttg gc gc ly A	cagaa gttta gaaaa cg gi la Va	agg agt aac at t al P	tacco cgagg acgat tc gt he Va gct Ala	ccattg yttaaa taatac ta ggt al Gly 15	1680 1740 1800 1849
gaad tgca atgg tatg aaad c at Me	aaagg gctct gggat cgtct cg ga et As l ata Ile	ctc Leu	gcaca tcaas gatct gccco gg ga rg G] ttg Leu 20	aacco gcgta gggg cccga g at Lu Me	ec aget to the control of the contro	tca tca cacgg ca go la A: tca Ser	aaggg tgèac gggac ca to la Se ccg Pro	g got c ato c gto cg to er C cac His 25	tgaaggetti ggtti gg tg ge gg ys G tat Tyr	ggat taca ttcc ga gg ly Gi 10 aag Lys	gccc tgtc tttc gc gc ly A: ctg Leu	ttc Phe	agg agt act tt al P ctc Leu 30	tacco cgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
gaad tgca atgg tatg aaad c· at Me ctg Leu	aaagg gctct gggat cgtct cg ga et As l ata Ile	geg g cet g ag g ac eg ac eg ctc Leu	gcaca tcaag gatet gccco gg ga rg G] ttg Leu 20	according to the second	ec aget accept All the Leu caa	tca tca tca tca tca Ser	aaggg tgèac gggac ca to la Se ccg Pro	g got c ato c gto cg to er C cac His 25	tgaaggetti ggtti ggtti gc gg ys G: tat Tyr	ggat taca ttcc ga gg ly G: 10 aag Lys	gccc tgtg tttg gc gc ly Al ctg Leu	gaga: gttt: gaaa; g ggl la V; ttc Phe	agg agt act tal P ctc Leu 30	tacco cgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
gaad tgca atgg tatg aaad c· at Me ctg Leu	aaagg gctct gggat cgtct cg ga et As l ata Ile	ctc Leu	gcaca tcaag gatet gccco gg ga rg G] ttg Leu 20	according to the second	ec aget accept All the Leu caa	tca tca tca tca tca Ser	aaggg tgcac gggac ca to la Se ccg Pro	g got c ato c gto cg to er C cac His 25	tgaaggetti ggtti ggtti gc gg ys G: tat Tyr	ggat taca ttcc ga gg ly G: 10 aag Lys	gccc tgtg tttg gc gc ly Al ctg Leu	ttc Phe gag Glu	agg agt act tal P ctc Leu 30	tacco cgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
gaad tgca atgg tatg aaad c· at Me ctg Leu	aaagg gctct gggat cgtct cg ga et As l ata Ile	geg g cet g ag g ac eg ac eg ctc Leu	gcaca tcaag gatet gccco gg ga rg G] ttg Leu 20	according to the second	ec aget accept All the Leu caa	tca tca tca tca tca Ser	aaggg tgèac gggac ca to la Se ccg Pro	g got c ato c gto cg to er C cac His 25	tgaaggetti ggtti ggtti gc gg ys G: tat Tyr	ggat taca ttcc ga gg ly G: 10 aag Lys	gccc tgtg tttg gc gc ly Al ctg Leu	gaga: gttt: gaaa; g ggl la V; ttc Phe	agg agt act tal P ctc Leu 30	tacco cgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
gaad tgca atgc tatc aaad c·at Me ctg Leu	aaagg gctct gggat cgtct cg ga et As l ata Ile ata Ile	tgg tgg tgg tgg tgg tgg tgg tgg tgg	gcaca tcaag gatet gccco gg ga rg Gl ttg Leu 20 tgg Trp	according to the control of the cont	cc agat to ge class con get Alice to Leu caa Gln	tca tca tca tca tca Ser tat Tyr	aaggg tgcac gggac ca to la Se ccg Pro ttt Phe 40	g gct c atç c g tç er C cac His 25 atc	tgaaggett ggttl ggtgg ys G tat Tyr acc Thr	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg	gcc tgtg tttg gc gc ly Al ctg Leu gcc Ala	ttc pagg gttta gaaaa gg gt la Va ttc Phe gag Glu 45	agg agt act tt tal P ctc Leu 30 gca Ala	taccoccgaggaggaggaggaggaggaggaggaggaggaggagga	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu	1680 1740 1800 1849 1897
gaad tgca atgc tatg aaad c at Me ctg Leu ctc	aaagg gctct gggat cgtct cg ga et As l ata Ile ata Ile	geg gee teet gee teet gee teet gee teet t	gcaca tcaag gatct gcccc gg ga rg G]  ttg Leu 20  tgg Trp	according to the coccording at the coccording at the coccording at the coccording to the coccordinate to the c	cc agat to get All to	tca ser tat tat tat	aaggg tgcac gggac ca tc la Se Pro ttt Phe 40	g gct c atç c gtç cg tç er C cac His 25 atc Ile	tgaaggetti getti getti ge ge ys G: tat Tyr acc Thr	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg	gcc tgtg tttg gc gc ly Al ctg Leu gcc Ala	ttc pagg gttta gaaaa gg gt la Va ttc Phe gag Glu 45	agg agt act tt al P ctc Leu 30 gca Ala	taccoccegages acgaint c githe Value of the V	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc	1680 1740 1800 1849
gaad tgca atgc tatg aaad c at Me ctg Leu ctc	aaagg gctct gggat cg ga et As l ata Ile ata Ile gtg Val	geg gee teet gee teet gee teet gee teet t	gcaca tcaag gatct gcccc gg ga rg G]  ttg Leu 20  tgg Trp	according to the coccording at the coccording at the coccording at the coccording to the coccordinate to the c	cc agat to get All to	tca Ser tat Tyr	aaggg tgcac gggac ca tc la Se Pro ttt Phe 40	g gct c atç c gtç cg tç er C cac His 25 atc Ile	tgaaggetti getti getti ge ge ys G: tat Tyr acc Thr	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg	gcc tgtg tttg gc gc ly Al ctg Leu gcc Ala	ttc pagg gttta gaaaa gg gt la Va ttc Phe gag Glu 45	agg agt act tt al P ctc Leu 30 gca Ala	taccoccegages acgaint c githe Value of the V	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc	1680 1740 1800 1849 1897
gaad tgca atgc tatg aaad c at Me ctg Leu ctc	aaagg gctct gggat cgtct cg ga et As l ata Ile ata Ile	geg gee teet gee teet gee teet gee teet t	gcaca tcaag gatct gcccc gg ga rg G]  ttg Leu 20  tgg Trp	according to the coccording at the coccording at the coccording at the coccording to the coccordinate to the c	cc agat to get All to	tca ser tat tat tat	aaggg tgcac gggac ca tc la Se Pro ttt Phe 40	g gct c atç c gtç cg tç er C cac His 25 atc Ile	tgaaggetti getti getti ge ge ys G: tat Tyr acc Thr	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg	gcc tgtg tttg gc gc ly Al ctg Leu gcc Ala	ttc pagg gttta gaaaa gg gt la Va ttc Phe gag Glu 45	agg agt act tt al P ctc Leu 30 gca Ala	taccoccegages acgaint c githe Value of the V	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc	1680 1740 1800 1849 1897
ctg ctau ctg ctau ctg Leu ctc	aaagg gctct gggat cg ga et As l ata Ile ata Ile yal 50	ctc ctc ctc teu tgg Trp 35	gcaca tcaac gatct gcccc gg ga rg G]  ttg Leu 20  tgg Trp  atc	acc gegggeccga ag at u Me acc Thr tta Leu	cc agat to aga co aga c	tca ser ttat ttat ttat ttat ttat ttat ttat tt	ca to la Se cog Pro ttt Phe 40 aac Asn	g got c ate c gte cg te er C cac His 25 atc Ile gtt Val	tgaaggettiggettige gg tat Tyr . acc Thr	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg	gcc gcly Ala ggc Gly 60	ttc gag tttc gaaa cg gt la Va ttc Phe gag Glu 45 cgc Arg	agg agt act tt al P ctc Leu 30 gca Ala	taccoccegaggacgate Value	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1680 1740 1800 1849 1897
ctg Leu ctc Leu caa Gln atc	aaagg gctct gggat cg ga et As l ata Ile ata Ile ytg Val 50	ctc ctc tgg Trp ctc	gcaca tcaac gatct gcccc gg ga rg Gl  ttg Leu 20  tgg Trp  atc Ile acg	acc grant grant grant grant acc Thr tta Leu	cca gcat to gc ct aa cc gg et Al ttg Leu caa Gln ccc Pro	tca ser tat tca ser tat Tyr ctc Leu 55	cac	c atomic cac His atomic structured at the Structure of th	tgaaggettiggttige gg ys Given Tyr acc Thr cgg Arg	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg ggg	gcc gcly Ala ggc Gly 60 atc	ttc gag tttc gaaa gg la V ttc Phe gag Glu 45 cgc Arg	agg agt aac tt t al P ctc Leu 30 gca Ala gat Asp	taccoccegaggaegaegaegaegaegaegaegaegaegaegaegae	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc	1680 1740 1800 1849 1897
ctg Leu ctc Leu caa Gln atc	aaagg gctct gggat cg ga et As l ata Ile ata Ile yal 50	ctc ctc tgg Trp ctc	gcaca tcaac gatct gcccc gg ga rg Gl  ttg Leu 20  tgg Trp  atc Ile acg	acc grant grant grant grant acc Thr tta Leu	cca gcat to gc ct aa cc gg et Al ttg Leu caa Gln ccc Pro	tca ser tat tca ser tat Tyr ctc Leu 55	cac	c atcomplete cac His atcomplete val	tgaaggettiggttige gg ys Given Tyr acc Thr cgg Arg	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg ggg	gcc gcly Ala ggc Gly 60 atc	ttc gag tttc gaaa gg la V ttc Phe gag Glu 45 cgc Arg	agg agt aac tt t al P ctc Leu 30 gca Ala gat Asp	taccoccegaggaegaegaegaegaegaegaegaegaegaegaegae	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc	1680 1740 1800 1849 1897 1945
ctg Leu ctc Leu caa Gln atc	aaagg gctct gggat cg ga et As l ata Ile ata Ile ytg Val 50	ctc ctc tgg Trp ctc	gcaca tcaac gatct gcccc gg ga rg Gl  ttg Leu 20  tgg Trp  atc Ile acg	acc grant grant grant grant acc Thr tta Leu	cca gcat to gc ct aa cc gg et Al ttg Leu caa Gln ccc Pro	tca ser tat tca ser tat Tyr ctc Leu 55	cac	c atcomplete cac His atcomplete val	tgaaggettiggttige gg ys Given Tyr acc Thr cgg Arg	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg ggg	gcc gcly Ala ggc Gly 60 atc	ttc gag tttc gaaa gg la V ttc Phe gag Glu 45 cgc Arg	agg agt aac tt t al P ctc Leu 30 gca Ala gat Asp	taccoccyange acgarded to get Ala cac His gcc Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc	1680 1740 1800 1849 1897 1945
ctg Leu ctc Leu caa Gln atc	aaagg gctct gggat cg ga et As l ata Ile ata Ile ytg Val 50	ctc ctc tgg Trp ctc	gcaca tcaac gatct gcccc gg ga rg Gl  ttg Leu 20  tgg Trp  atc Ile acg	acc grant grant grant grant acc Thr tta Leu	cca ggc ttg Leu caa Gln ccc Pro	tca ser tat tca ser tat Tyr ctc Leu 55	cac	c atcomplete cac His atcomplete val	tgaaggettiggttige gg ys Given Tyr acc Thr cgg Arg	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg ggg Gly cta Leu	gcc gcly Ala ggc Gly 60 atc	ttc gag tttc gaaa gg la V ttc Phe gag Glu 45 cgc Arg	agg agt aac tt t al P ctc Leu 30 gca Ala gat Asp	taccoccyange acgarded to get Ala cac His gcc Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr	1680 1740 1800 1849 1897 1945
ctc Leu caa Gln atc Ile 65	aaagggctctgggatcggat As I ata Ile stg Val ctc Leu atc	ctc	tcaacgatctgccccgg gatctgccccgg gatctgg gatctgg Leu 20 tgg Trp atc Ile acg Thr	acc geggg geggg geggg ag at u Me acc Thr tta Leu ccc Pro	cc agat to get A. S. Leu caa Gln ccc Pro gcg Ala 70 ata	tca ser tat tca ser tat Tyr ctc Leu 55 atc	aagggctgcacacacacacacacacacacacacacacaca	cac gtc cac His atc Ile gtt Val cca Pro	tgaaggettiggettige gggys G: tat Tyr acc Thr cgg Arg gag Glu ctc	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg	gcc gcly Ala gcc Ala ggc Gly 60 atc Ile gtg	ttc gagg ttt: gaaa gg gg la V ttc Phe gag Glu 45 cgc Arg ttt Phe	agg agg agg agg ct t t t t l l l l l l l l l l l l l l	taccoccegages acgaint to get Ala cac His gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1680 1740 1800 1849 1897 1945
ctc Leu caa Gln atc Ile 65	aaagg gctct gggat cg ga et As l ata Ile ata Ile yal 50 ctc	ctc	tcaacgatctgccccgg gatctgccccgg gatctgg gatctgg Leu 20 tgg Trp atc Ile acg Thr	acc geggg geggg geggg ag at u Me acc Thr tta Leu ccc Pro	cc agat to get A. S. Leu caa Gln ccc Pro gcg Ala 70 ata	tca ser tat tca ser tat Tyr ctc Leu 55 atc	aagggctgcacacacacacacacacacacacacacacaca	cac gtc cac His atc Ile gtt Val cca Pro	tgaaggettiggettige gggys G: tat Tyr acc Thr cgg Arg gag Glu ctc	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg	gcc gcly Ala gcc Ala ggc Gly 60 atc Ile gtg	ttc gagg ttt: gaaa gg gg la V ttc Phe gag Glu 45 cgc Arg ttt Phe	agg agg agg agg ct t t t t l l l l l l l l l l l l l l	taccoccegages acgaint to get Ala cac His gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1680 1740 1800 1849 1897 1945 1993
ctc Leu caa Gln atc Ile 65	aaagggctctgggatcggat As I ata Ile stg Val ctc Leu atc	ctc	tcaacgatctgccccgg gatctgccccgg gatctgg gatctgg Leu 20 tgg Trp atc Ile acg Thr	acc geggg geggg geggg ag at u Me acc Thr tta Leu ccc Pro	cc agat to get A. S. Leu caa Gln ccc Pro gcg Ala 70 ata	tca ser tat tca ser tat Tyr ctc Leu 55 atc	aagggctgcacacacacacacacacacacacacacacaca	cac gtc cac His atc Ile gtt Val cca Pro	tgaaggettiggettige gggys G: tat Tyr acc Thr cgg Arg gag Glu ctc	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg	gcc gcly Ala gcc Ala ggc Gly 60 atc Ile gtg	ttc gagg ttt: gaaa gg gg la V ttc Phe gag Glu 45 cgc Arg ttt Phe	agg agg agg agg ct t t t t l l l l l l l l l l l l l l	taccoccegages acgaint to get Ala cac His gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1680 1740 1800 1849 1897 1945 1993
ctc Leu caa Gln atc Ile 65	aaagggctctgggatcggat As I ata Ile stg Val ctc Leu atc	ctc	tcaacgatctgccccgg gatctgccccgg gatctgg gatctgg Leu 20 tgg Trp atc Ile acg Thr	acc geggg geggg cegga ag at u Me acc Thr tta Leu ccc Pro tgc Cys	cc agat to get A. S. Leu caa Gln ccc Pro gcg Ala 70 ata	tca ser tat tca ser tat Tyr ctc Leu 55 atc	aagggctgcacacacacacacacacacacacacacacaca	cac gtc cac His atc Ile gtt Val cca Pro	tgaag getti ggtti ge gg ys G: tat Tyr acc Thr cgg Arg gag Glu ctc Leu	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg	gcc gcly Ala gcc Ala ggc Gly 60 atc Ile gtg	ttc gagg ttt: gaaa gg gg la V ttc Phe gag Glu 45 cgc Arg ttt Phe	agg agg agg agg ct t t t t l l l l l l l l l l l l l l	taccoccegages acgains to get he Vala cac His gcc Ala atc Ile gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1680 1740 1800 1849 1897 1945 1993
ctc Leu caa Gln atc Ile 65	aaagggctctgggatcggat As I ata Ile stg Val ctc Leu atc	ctc	ttg Leu 20 tgg Trp atc Ile acg Thr	acc geggg geggg geggg ag at u Me acc Thr tta Leu ccc Pro tgc Cys	cc agat to ge as a co ge at A. S. Leu caa Gln ccc Pro gcg Ala 70 ata Ile	tca ser tat tca ser tat Tyr ctc Leu ctc Leu	caggrade a see coggrade a see	cac gtc cac His atc Ile gtt Val cca Pro	tgaag getti getti ge gg ys G: tat Tyr acc Thr cgg Arg gag Glu ctc Leu	ggat taca ttcc ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg	gcc gcly Ala gcc Ala ggc 60 atc Ile gtg Val	ttc gagg ttt: gaaa gg gg la Va ttc Phe gag Glu 45 cgc Arg ttt Phe	agg agg agg agg agg ct t t tal P ctc Leu 30 gca Ala gat Asp acc Thr	taccoccegages acgainst the Value of Val	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1680 1740 1800 1849 1897 1945 1993
ctg Leu ctc Leu caa Gln atc Ile 65 aaa Lys	adagggctctgggtctgggt Asil at a Ile at a Ile store the Ile acc	ctc (ctc) (c	gcaca tcaac gatct gcccc gg ga rg Gl  ttg Leu 20 tgg Trp atc Ile acg Thr ctc Leu gtg	acc gradia de la companya de la comp	cc agat to ge at a co get A:  ttg caa Gln ccc Pro gcg Ala 70 ata Ile	tca ser tat tca ser tat Tyr ctc Leu 55 atc Ile ctc Leu	ccg Pro ttt Phe 40 aac Asn cac His	cac gtcg cac His 25 atc Ile gtt Cac Pro	tgaag getti ggtti ge gg ys G tat Tyr acc Thr cgg Glu ctc Leu 90	ggat taca taca ggag taca taca ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg Met cac	gcc gcly Ala ggc Ala	ttc paag tttc gaaa gg la Va ttc Phe gag Glu 45 cgc Arg ttt Phe	agg	taccoccyagos acgait to go the Vala cac His gcc Ala atc Ile gct Ala 95 cgt	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1680 1740 1800 1849 1897 1945 1993 2041
ctg Leu ctc Leu caa Gln atc Ile 65 aaa Lys	adagggctctgggtctgggt As I at a I le gtg V 50 ctc Leu at c	ctc (ctc) (c	gcaca tcaac gatct gcccc gg ga rg Gl  ttg Leu 20 tgg Trp atc Ile acg Thr ctc Leu gtg	acc gradia de la companya de la comp	cc agat to ge at a co get A:  ttg caa Gln ccc Pro gcg Ala 70 ata Ile	tca ser tat tca ser tat Tyr ctc Leu 55 atc Ile ctc Leu	ccg Pro ttt Phe 40 aac Asn cac His	cac gtcg cac His 25 atc Ile gtt Cac Pro	tgaag getti ggtti ge gg ys G tat Tyr acc Thr cgg Glu ctc Leu 90	ggat taca taca ggag taca taca ga gg ly G: 10 aag Lys agg Arg Gly cta Leu 75 atg Met cac	gcc gcly Ala ggc Ala	ttc paag tttc gaaa gg la Va ttc Phe gag Glu 45 cgc Arg ttt Phe	agg	taccoccyagos acgait to go the Vala cac His gcc Ala atc Ile gct Ala 95 cgt	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1680 1740 1800 1849 1897 1945 1993 2041

				cgg Arg										2185
	_	_	_	gcc Ala	_	_			_	-		_		2233
				gac Asp										2281
_	_	-		gtc Val 165	_			_	_		_	_		2329
		_	_	acc Thr			_		_		_		_	2377
				agg Arg										2425
	~		_	glå aaa		_					_	_		2473
				ggc Gly										2521
				cag Gln 245										2569
				ctg Leu										2617
				Gly										2665
				acc Thr										2713
				cgt Arg										2761
				acg Thr 325										2809

	_				agc Ser							_			_	2857
_			_		ggt Gly		_		_		_			_		2905
					gcc Ala											2953
_		_		_	gag Glu 390		_				_			_	_	3001
					tcc Ser											3049
gcc Ala	cat His	cta Leu	cac His 420	gcc Ala	cct Pro	act Thr	ggt Gly	agc Ser 425	ggc Gly	aag Lys	agc Ser	act Thr	aag Lys 430	gtg Val	ccg Pro	3097
					caa Gln											3145
					ggt Gly											3193
					aga Arg 470			_					_		_	3241
ccc Pro	atc Ile	acg Thr	tac Tyr	tcc Ser 485	acc Thr	tat Tyr	ggc Gly	aag Lys	ttt Phe 490	ctt Leu	gcc Ala	gac Asp	ggt Gly	ggt Gly 495	tgc Cys	3289
					gac Asp											3337
gac Asp					ctg Leu											3385
					ctc Leu											3433
tcg Ser 545	gtc Val	acc Thr	gtg Val	cca Pro	cat His 550	cca Pro	aac Asn	atc Ile	gag Glu	gag Glu 555	gtg Val	gct Ala	ctg Leu	tcc Ser	agc Ser 560	3481

	gga Gly									3529
_	eja aaa	 					-	_	_	3577
	ctc Leu									3625
	cgg Arg 610									3673
	gta Val									3721
	gtg Val									3769
	gac Asp									3817
	tca Ser									3865
	tac Tyr 690									3913
	tcg Ser	 _	_	 _	-	 				 3961
	acg Thr									4009
	GJÀ aaa									4057
	aca Thr									4105
	gca Ala 770									4153

# **53 / 93** .

_	_		_	_	_	cca Pro			_		_					4201
_					_	cct Pro	_	_				_		_	-	4249
						caa Gln										4297
				_	_	tgc Cys	_	_	_	_	_					4345
agc Ser	acc Thr 850	tgg Trp	gtg Val	ctg Leu	gta Val	ggc Gly 855	gga Gly	gtc Val	cta Leu	gca Ala	gct Ala 860	ctg Leu	gcc Ala	gcg Ala	tat Tyr	4393
						gtg Val								Leu		4441
						ccc Pro										4489
						gcc Ala										4537
						ttc Phe									caa Gln	4585
						gag Glu 935										4633
						ttc Phe										4681
						gca Ala										4729
						gca Ala										4777
			His			ctg Leu		Asn					Trp			4825

Ala Gln Leu 1010		cc agc gct ro Ser Ala 1015				
atc gct gga Ile Ala Gly 1025	Ala Ala Va		Ile Gly			
gat att ttg Asp Ile Leu				Ala Gly Al		Ala
ttt aag gtc Phe Lys Val						
cta ctc cct Leu Leu Pro 107	Ala Ile L		Gly Ala	Leu Val Va		
tgc gca gcg Cys Ala Ala 1090						
cag tgg atg Gln Trp Met 1105	Asn Arg Le		Phe Ala			•
tcc ccc acg Ser Pro Thr				Ala Ala Al		Thr
	His Tyr Va 1125 tct agt ct	al Pro Glu	Ser Asp 1130 act cag	Ala Ala Al ctg ctg aa	a Arg Val 113 g agg ctt	Thr 5 cac 5257
Ser Pro Thr	His Tyr Va 1125 tct agt co Ser Ser Le 1140 aac gag ga Asn Glu As	al Pro Glu  tt acc atc eu Thr Ile ac tgc tcc	Ser Asp 1130 act cag Thr Gln 1145 acg cca Thr Pro	Ala Ala Al ctg ctg aa Leu Leu Ly tgc tcc gg Cys Ser Gl	a Arg Val 113: ag agg ctt rs Arg Leu 1150 gc tcg tgg	Thr 5 cac 5257 His cta 5305
cag atc ctc Gln Ile Leu cag tgg atc Gln Trp Ile	tct agt conserved to the ser Ser Let 1140 aac gag ga Asn Glu As to tgg gat to	al Pro Glu  tt acc atc eu Thr Ile  ac tgc tcc sp Cys Ser 1160 gg ata tgc	act cag Thr Gln 1145 acg cca Thr Pro acg gtg	ctg ctg aa Leu Leu Ly tgc tcc gg Cys Ser Gl 11 ttg act ga	a Arg Val 113: ag agg ctt rs Arg Leu 1150 ac tcg tgg ry Ser Trp 65	Thr 5 5 5257 His 5305 Leu acc 5353
cag atc ctc Gln Ile Leu  cag tgg atc Gln Trp Ile 115  aga gat gtt Arg Asp Val	tct agt construction of the series of the se	tt acc atc eu Thr Ile ac tgc tcc sp Cys Ser 1160 gg ata tgc rp Ile Cys 1175 tc ctg ccg	act cag Thr Gln 1145  acg cca Thr Pro  acg gtg Thr Val  cga ttg Arg Leu	ctg ctg aa Leu Leu Ly  tgc tcc gg Cys Ser Gl 11  ttg act ga Leu Thr As 1180  ccg gga gt	a Arg Val 113  ag agg ctt rs Arg Leu 1150  ac tcg tgg ay Ser Trp 65  at ttc aag ap Phe Lys	Thr 5
cag atc ctc Gln Ile Leu  cag tgg atc Gln Trp Ile 115 aga gat gtt Arg Asp Val 1170  tgg ctc cag Trp Leu Gln	His Tyr Variation 1125  tot agt of Ser Ser Let 1140  aac gag ga Asn Glu As 5  tgg gat tg Trp Asp Tr  tcc aag cr Ser Lys Let 1125	tt acc atc eu Thr Ile ac tgc tcc sp Cys Ser 1160 gg ata tgc rp Ile Cys 1175 tc ctg ccg eu Leu Pro 190 ac aag gga	act cag Thr Gln 1145 acg cca Thr Pro acg gtg Thr Val cga ttg Arg Leu	ctg ctg aa Leu Leu Ly  tgc tcc gg Cys Ser Gl 11  ttg act ga Leu Thr As 1180  ccg gga gt Pro Gly Va 1195  cgg ggc ga Arg Gly As	a Arg Val 113  ag agg ctt rs Arg Leu 1150  ac tcg tgg ry Ser Trp 65  at ttc aag rp Phe Lys  ac ccc ttc al Pro Phe ac ggc atc	Thr 5

	agg atc gtg ggg Arg Ile Val Gly			
	ecc att aac geg Pro Ile Asn Ala 1255	Tyr Thr Thr		
	aat tat tot agg Asn Tyr Ser Arg 1270			
	ott acg cgg gtg Val Thr Arg Val 1285		His Tyr Val T	
Thr Thr Asp A	ac gta aag tgc Asn Val Lys Cys 1300		Val Pro Ala P	
	gtg gat ggg gtg Val Asp Gly Val			
	cta cgg gag gag Leu Arg Glu Glu 1335	Val Thr Phe		
	ggg tca cag ctc Gly Ser Gln Leu 1350			
	cc atg ctc acc Ser Met Leu Thr 1365		His Ile Thr A	
Ala Lys Arg A	agg ctg gcc agg Arg Leu Ala Arg .380		Pro Ser Leu A	
	eag ctg tct gcg Eln Leu Ser Ala			
•	ccc ccg gac gct Ser Pro Asp Ala 1415	Asp Leu Ile		
	atg ggc ggg aac Met Gly Gly Asn 1430			
gta gta att t Val Val Ile L	tg gac tct ttc Seu Asp Ser Phe 1445	gag ccg ctc Glu Pro Leu 1450	Gln Ala Glu G	ag gat gag 6169 lu Asp Glu 1455

agg gaa gta tcc Arg Glu Val Ser 1460	Val Pro Ala				217
cct ega geg atg Pro Arg Ala Met 1475			_	Pro Pro Leu	65
tta gag tcc tgg Leu Glu Ser Trp 1490		Asp Tyr Val			313
tgt cca ctg ccg Cys Pro Leu Pro 1505		-			861
aag agg acg gtt Lys Arg Thr Val		_	Val Ser Ser		109
gag ctc gcc aca Glu Leu Ala Thr 1540	Lys Thr Phe				157
agc ggc acg gca Ser Gly Thr Ala 1555				Asp Gly Asp	505
gcg gga tcc gac Ala Gly Ser Asp 1570		Tyr Ser Ser	•		553
gag ccg ggg gat Glu Pro Gly Asp 1585					01
gag gag gct agt Glu Glu Ala Ser			Ser Met Ser		49
aca ggc gcc ctg Thr Gly Ala Leu 1620	Ile Thr Pro			-	97
atc aat gċa ctg Ile Asn Ala Leu 1635				Leu Val Tyr	45
gct aca aca tct Ala Thr Thr Ser 1650		Ser Leu Arg			93
gac aga ctg cag Asp Arg Leu Gln 1665					41

				Ser					Lys			tcc Ser		Glu	6889
			Leu					Ser				aaa Lys 1710	Phe		6937
		Lys					Leu					gtt Val			6985
-	Ser			_	_	Leu	_	_	_		Glu	aca Thr			7033
Thr					ьуз					Суз		caa Gln			7081
				Pro					Val			gat Asp		Gly	7129
			Glu					$\mathtt{Tyr}$				tcc Ser 1790	Thr		7177
		Val					${\tt Tyr}$					tct Ser			7225
	Val					Asn					Lys	aaa Lys			7273
Gly					Thr					Ser		gtc Val			7321
_		_	_	Glu					Gln	_	_	gac Asp	_	Ala	7369
			Gln					Leu				ctt Leu 1870	Tyr		7417
 		Leu					Gly			_		tat Tyr 5	_		7465
	Ala					Thr					Asn	acc Thr			7513

_	Tyr	_	_	_		Ala	_	_	_	_	Ala	_	ctc Leu	_	_	7561
					Cys					Val			tgt Cys		Ser	7609
				Glu					Leu				acg Thr 195	Glu		7657
_		_	Tyr		-			Gly	-	_			cca Pro 5	_		7705
		Glu					Cys					Ser	gtc Val			7753
	Ala					Val					Arg		ccc Pro			7801
					Ala					Arg			cca Pro		Asn	7849
				Asn					Ala				tgg Trp 2030	Āla		7897
			Met					Ser					cag Gln			7945
		Lys					Gln					Сув	tac Tyr			7993
	Pro					Gln					Leu		ggc			8041
					Ser					Glu			agg Arg		Ala	8089
tca Ser	tgc Cys	ctc Leu	agg Arg 2100	Lys	ctt Leu	gjå aaa	gta Val	ccg Pro 2105	Pro	ttg Leu	cga Arg	gtc Val	tgg Trp 2110	Arg	cat His	8137
			Ser					Leu					gly 333			8185

Ala Thr Cys Gly Lys Tyr Leu Phe 2130 2135	20 2 2 25 -	8233
aaa ctc act cca atc ccg gct gcg Lys Leu Thr Pro Ile Pro Ala Ala 2145 2150		8281
ttc gtt gct ggt tac agc.ggg gga Phe Val Ala Gly Tyr Ser Gly Gly 2165		8329
gcc cga ccc cgc tgg ttc atg tgg Ala Arg Pro Arg Trp Phe Met Trp 2180	5 555	8377
gta ggc atc tat cta ctc ccc aac Val Gly Ile Tyr Leu Leu Pro Asn 2195 2200	Arg *	8427
	ttttttttt tttttttt ttttttt {	
	: ttttttttcc ttttctttcc tttggtggct {   aaaggtccgt gagccgcttg actgcagaga {	
gtgctgatac tggcctctct gcagatcaag		8638
<210> 7 <211> 8638 <212> DNA <213> HCV		
<220>		
<221> CDS		
<221> CDS <222> (1802)(8407) <400> 7		<b>.</b> 0
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407) &lt;400&gt; 7 gccagccccc gattgggggc gacactccacc</pre>	catagatcac teceetgtga ggaactaetg 6	
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407) &lt;400&gt; 7 gccagccccc gattgggggc gacactccactcttcacgca gaaagcgtct agccattggggg</pre>	catagatcac teceetgtga ggaactaetg e ttagtatgag tgtegtgeag eeteeaggae l eggaaceggt gagtacaceg gaattgeeag l	120
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg</pre>	ttagtatgag tgtcgtgcag cctccaggac 1	120 180
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagegtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc</pre>	ttagtatgag tgtcgtgcag cctccaggac l cggaaccggt gagtacaccg gaattgccag l ctcaatgcct ggagatttgg gcgtgccccc 2 gcgaaaggcc ttgtggtact gcctgatagg 3	120 180 240 300
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccactctcccgaaaagegtct agccatggcgccccccccccccccccccccccccc</pre>	ttagtatgag tgtcgtgcag cctccaggac I cggaaccggt gagtacaccg gaattgccag I ctcaatgcct ggagatttgg gcgtgcccc 2 gcgaaaggcc ttgtggtact gcctgatagg 3 gaccgtgcac catgagcacg aatcctaaac 3	120 180 240 300 360
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccact tcttcacgca gaaagegtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga</pre>	ttagtatgag tgtcgtgcag cctccaggac I cggaaccggt gagtacaccg gaattgccag I ctcaatgcct ggagatttgg gcgtgcccc 2 gcgaaaggcc ttgtggtact gcctgatagg 3 gaccgtgcac catgagcacg aatcctaaac 3 ttgaacaaga tggattgcac gcaggttctc 4	120 180 240 300 360 420
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccactctcccga gaaagcgtct agccatggcgccccccctccc gggagagcca tagtggtctggacgaccggg tcctttcttg gatcaacccgggcgagactgc tagccgagta gtgttgggtcggtgttgctggtgttgcgagactgc gtgccccggg aggtctcgtactaaagaaa aaccaaaggg cgcgccatgacggccgcttg ggtggagagg ctattcggct</pre>	ttagtatgag tgtcgtgcag cctccaggac 1 cggaaccggt gagtacaccg gaattgccag 1 ctcaatgcct ggagatttgg gcgtgcccc 2 gcgaaaggcc ttgtggtact gcctgatagg 3 gaccgtgcac catgagcacg aatcctaaac 3 ttgaacaaga tggattgcac gcaggttctc 4 atgactggc acaacagaca atcggctgct 4	120 180 240 300 360 420
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccact tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga cggccgcttg ggtggagagg ctattcggcc ctgatgccg cgtgttccgg ctgtcagcgc</pre>	ttagtatgag tgtcgtgcag cctccaggac g cggaaccggt gagtacaccg gaattgccag g ctcaatgcct ggagatttgg gcgtgcccc 2 gcgaaaggcc ttgtggtact gcctgatagg 3 gaccgtgcac catgagcacg aatcctaaac 3 ttgaacaaga tggattgcac gcaggttctc 4 atgactggc acaacagaca atcggctgct 4 aggggcgccc ggttctttt gtcaagaccg 5	120 180 240 300 360 420 480
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgcatga cggccgcttg ggtggagagg ctattcggc ctgatgccgc cgtgttccgg ctgtcagcg acctgtccgg tgccctgaat gaactgcagg cgacggcgt tccttgcgca gctgtgctcg</pre>	ttagtatgag tgtcgtgcag cctccaggac geggaaccggt gagtacaccg gaattgccag getccaccg gagtatgccac geggaaccgct tgtggtact gcctgatagg geggacaccg catgagcacg aatcctaaac geaccgtgcac catgagcacg aatcctaaac geaggttctc atgactggc acaacagaca atcggctgct aggggcgccc ggttctttt gtcaagaccg gacggtctct gacgaggcagc gcggctatcg tggctggca acggttctc gacgaggcagc gcggctatcg tggctggca acggttctc	120 180 240 300 360 420 480 540 660
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccacc tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccaggta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgcatga cggccgcttg ggtggagagg ctattcggc ctgatgccgc cgtgttccgg ctgtcagcgc acctgtccgg tgccctgaat gaactgcagg cgacggcgt tccttgcgca gctgtgctcg tgctattggg cgaagtgccg gggcaggatc</pre>	ttagtatgag tgtcgtgcag cctccaggac geggaaccggt gagtacaccg gaattgccag getccaaggac cccaatgaccg gagtattgg gcgtgcccc gegaaaggcc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac geaggttctc gagggcacagacagacagacagacagacagacagacag	120 180 240 300 360 420 480 540 660 720
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccacc tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccaggta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgcatga cggccgcttg ggtggagagg ctattcggc ctgatgccgc cgtgttccgg ctgtcagcg acctgtccgg tgccctgaat gaactgcagg cgacggcgt tccttgcgca gctgtgctcg tgctattggg cgaagtgccg gggcaggatc aggtatccat catggctgat gcaatgcggc</pre>	ttagtatgag tgtcgtgcag cctccaggac gcggaaccggt gagtacaccg gaattgccag gctccaatgcc ggagatttgg gcgtgcccc gggaaaggcc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg aatcctaaac gcaggttctc gaggggcgcc ggttctttt gtcaagaccg gacggcacc ggttctttt gtcaagaccg gacggcacc gcggctatcg tggctgcca gacgttgtcac tgaagcggga agggactggc ggctgcatc tcaccttgct cctgccgaga gggctgcatac gcggctatacc gctgccgaga gggctgcatac gcggctatacc gctgccgaga gggctgcatac gctgccatac gcttgatccc gctacctgcc	120 180 240 300 360 420 480 540 660 720
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaa aaccaaaggg cggcgcttg ggtgagagg ctgatgccgc cgtgttccgg ctattcggc acctgtccgg tgcctgaat gaactgcagg cgacggggt tccttgcga gcggcaggatc gacgggcgt tccttgcgca gcggcaggatc aggtatccat catggctgat gcaatgcggc cattcgacca ccaagcgaaa catcgcatcg</pre>	ttagtatgag tgtcgtgcag cctccaggac geggaaccggt gagtacaccg gaattgccag geggaaccggt gagtacaccg gaattgccag geggaaccgc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg aatcctaaac gacgggttctc atgaccggg acaacagaca atcggctgct agggggcgccc ggttcttttt gtcaagaccg gacggtatcg tggctgcac tgaagcggga agggactggc acgttgtcac tgaagcggga agggactggc tcctgtcatc tcaccttgct cctgccgaga gggctgcatac gagcggcacg tactcggatg gaagccggtc aggcagaccg tactcggatg gaagccggtc aggcagaccg tactcggatg gaagccggtc	120 180 240 300 360 420 480 540 660 720 780 840
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta cccaaagaa aaccaaaggg cggccatga cggccgcttg ggtgagagg ctattcggc ctgatgccgc cgtgttccgg ctgtcagcg cgacggtgt tccttgcgca gactgcagg cgacggggt tccttgcgca gcggcaggatc tgctattggg cgaagtgccg gggcaggatc aagtatccat catggctgat gcaatgcggc cattcgacca ccaagcgaaa catcgcatcg ttgtcgatca ggatgatctg gacgaagaga</pre>	ttagtatgag tgtcgtgcag cctccaggac geggaaccggt gagtacaccg gaattgccag geggaaccggt gagtacaccg gaattgccag geggaaccgc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg aatcctaaac gaggggcgcc ggttctttt gtcaagaccg gagggtctct gacgaggcacc ggttcttttt gtcaagaccg gacggctatcg tggctgcac tgaagcggga agggactggc ggctgcatc tcaccttgct cctgccgaga gggctgcatac gagcggcacg tactcggatg gaagccggtc gaccaggcc gaactgttcg atcagggcac gaactgttcg gaagcacg cgcgccagcc gaactgttcg	120 180 240 300 360 420 480 560 720 780 840 900
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccggga gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga cggcgcttg ggtgagagag ctattcggct ctgatgccgc cgtgttccgg ctgtcagcg acctgtccgg tgccctgaat gaactgcagg cgacggggt tccttgcgca gctgtcagg ctgatgccgc cgtgttccgg gtgcaggatc gacgggcgt tccttgcgca gctgtgctcg gcgacgggct tccttgcgca gctgtgctcg tgctattggg cgaagtgccg gggcaggatc aagtatccat catggctgat gcaatgcggc cattcgacca ccaagcgaaa catcgcatcg gctggctcaa ggcgccatg gccgacggcg gcttgccga tatcatggtg gaaaatggcc gcttgccga tatcatggtg gaaaatggcc gcttgccga tatcatggtg gaaaatggcc</pre>	ttagtatgag tgtcgtgcag cctccaggac cggaaccggt gagtacaccg gaattgccag cctcaatgcct gagtacaccg gaattgccag cctcaatgcct gagtacaccg gagtattgg gcgtgccccc gggaaaggcc ttgtggtact gcctgatagg accgtgcac catgagcacg aatcctaaac gaggggcgcc ggttcttttt gtcaagaccg acgggttctc gaggggcacc ggttcttttt gtcaagaccg acgggtatcg tggctgcac tgaagcggga agggactggc acggttgctc cctgccatc tcaccttgct cctgccgaga ggctgcatac gcttgatccg gcaacctgcc gaaccggcc gaacctgcc gaggatccgt cgcgccagcc gaactgttcg gaggatcccgt cgtgacccat ggcgatgcct ggctttttctg attcatcgac tgtggccagcc gaactgttcg gaggatctcgt cgtgacccat ggcgatgcct ggcgctttttttttt	120 180 240 300 360 420 480 540 660 720 780 840 990 960 1020
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga cggcgcttg ggtgagagag ctattcggct ctgatgccgc cgtgttccgg ctgtcagcg acctgtcgg tgccctgaat gaactgcagg gcgacggt tccttgcgca gctgtcagg cgacggggt tccttgcga gaactgcagg cgacgggcgt tccttgcgca gctgtgctcg acctgtccgg cgaagtgccg gggcaggatc aagtatccat catggctgat gcaatgcgg cattcgaca ccaagcgaaa catcgcatcg gctgcctaa ggcgccatg gacgagagg ccaggctcaa ggcgccatg gctggcgt tcctgggg aagtatcaa catggctgat cattcgacca ccaagcgaaa catcgcatcg gctgccgaa tatcatggtg gacagacatag gctggcgtgggggggcgcatg ccggacggcg gcttgccgaa tatcatggtg tgggtgtggc ggaccgctat caggacatag</pre>	ttagtatgag tgtcgtgcag cctccaggac gagtacaccg gagtacaccg gaattgccag gctcaatgccag gagtattgg gcgtgccccc gagacaggcc ttgtggtact gcctgatagg gagccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg aatcctaaac gaggggcgcc ggttctttt gtcaagaccg gaggttctc gaggggcgcc ggttctttt gtcaagaccg acgaggcacc tgaagcggaa agggactggc gactgtcac tgaagcggga agggactggc ggctgcatac gctgaccg gctgataccg gcgctacct gagcgagcacg tactcggat gaagccggtc agggaccac gacctgtcc gaggatctcgt cgtgaccac ggcgatgcct ggcttttctgg attcatcgac tgtggccagc ggcttttctgg attcatcgac tgtggccagc gcttgaccac gcttgaccat ggcgatgcct gcttggctac ccgtgatatt gctgaagagc	120 180 240 300 360 420 480 560 720 780 840 990 960 1020
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga cggcgcttg ggtgagagg ctattcggct ctgatgccgc cgtgttccgg gtgcaggcg acctgtcgg tgccctgaat gaactgcagg cgacggggt tccttgcgca gctgtgcgcg acctgtcgg ggtgagagag ccattcggc cgtgttccgg ggcaggatc tgctattggg cgaagtgccg ggcaggatc aagtatccat catggctgat gcaatgcggc cattcgacca ccaagcgaaa catcgcatcg ccaggctcaa ggcgccatg gacgagagc gcttgccgaa tatcatggtg tgggtgtggc ggaccgctat caggacatag tgggtgtggc atgggctgac cgcttcctcg</pre>	ttagtatgag tgtcgtgcag cctccaggac gagtacccgg gagtacaccg gaattgccag gcctcaatgcc gagtacaccg gaattgccag gcctcaatgcc gagtattgg gcgtgccccc gcgaaaggcc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg aatcctaaac gcaggttctc gagggcgcc ggttctttt gtcaagaccg gaggttctttt gtcaagaccg gaggttctttt gtcaagaccg gaggttctttt gtcaagaccg gaggcatcgc tggaggcac gctgatacc gcgtgcaacc gctgatacc gctgccagaa ggcgctaccg gagctgcacc gagcgcaccg gaactgtccg gagcaccg tactcggat gaagccggtc ggttcttttt gtcacttgcc gaaccgtcacc gctgccagcc gaactgtccg gagcacccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat gctgtgccac ccgtgatatt gctgaagagc gcgttttaccg tactcgccgc ccgattcgc tactcgccgc cccgattcgc tactcgccat cccgattcgc tactcgcat cccgattcgc tactcgcat cccgattcgc tactcgcat cccgattcgc tactcgcat cccgattcgcat cccgattcgcat cccgattcgcat cccgattcgcat cccgattcgc	120 180 240 300 360 420 480 560 720 780 840 990 960 1080 1140
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccac tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga cggcgcttg ggtgagagg ctattcggct ctgatgccgc cgtgttccgg ctgtcggcg acctgtccgg tgccctgaat gaactgcagg cgacggggt tccttgcgca gcggcaggatc tgctattggg cgaagtgccg ggcaggatc tgctattggg cgaagtgccg ggcaggatc agtatccat catggctgat gcaatgcggc cattcgacca ccaagcgaaa catcgcatcg ccaggctcaa ggcgcatg cccgacggcg gcttgccgaa tatcatggtg tgggtgtggc ggaccgctat caggacatag ttggcggca atgggctgac cgcttcctcg agcgcatcgc cttctatcgc cttcttgacg agcgcatcgc</pre>	ttagtatgag tgtcgtgcag cctccaggac gagtacaccg gagtacaccg gaattgccag gagtacaccg gaattgccag gctcaatgccag gagtattgg gcgtgccccc gagagatttgg gcgtgccccc gagacaggcc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg acaccagaca atcggctgct gagggcacg ggttctttt gtcaagaccg gcggctattg tggctggca agggacagc gcggctatcg tggctgcac catgagcagga agggactggc gactgatac gctgataccg gcagcagca gctgataccg gaagcagga gagcaggac gactgatcg gaagcagga tactcggatgaaccg gaactgttcg gcttttctgg attcatcgac tgtggccac ggtttttctgg attcatcgac tgtggccagc gatttttctgg attcatcgac gctgatatt gctgaagagc gagttcttacg gattttactg agttcttctg agttcgccc cagatttaa gattcttctg agttcgccc cagatttaa gattcttctg agttcgccc cagatttaa gattcttctg agttcgcc cagatgttaa	120 180 240 300 360 420 480 540 660 720 780 840 9900 1020 1140
<pre>&lt;221&gt; CDS &lt;222&gt; (1802)(8407)  &lt;400&gt; 7 gccagccccc gattgggggc gacactccacc tcttcacgca gaaagcgtct agccatggcg ccccctccc gggagagcca tagtggtctg gacgaccggg tcctttcttg gatcaacccg gcgagactgc tagccgagta gtgttgggtc gtgcttgcga gtgccccggg aggtctcgta ctcaaagaaa aaccaaaggg cgcgccatga cggccgcttg ggtgagagag ctattcggct ctgatgccgc cgtgttccgg ctgtcaggcg acctgtccgg tgccctgaat gaactgcagg cgacggggt tccttgcgca gctgtgctcg tgctattggg cgaagtgccg ggcaggatc acttcaca catggctgat gcaatgcggc tattcgacca catggctgat gcaatgcggc cattcgacca caagcgaaa catcgcatcg ccaggctcaa ggcgcatg gacaaaggc gcttgccgaa tatcatggtg tgggtgtggc ggaccgctat caggacatag ttggcggcaa atgggctgac cgcttcctcg agcgcacca cggtttccct ctagcgggat</pre>	ttagtatgag tgtcgtgcag cctccaggac gagtacccgg gagtacaccg gaattgccag gcctcaatgcc gagtacaccg gaattgccag gcctcaatgcc gagtattgg gcgtgccccc gcgaaaggcc ttgtggtact gcctgatagg gaccgtgcac catgagcacg aatcctaaac gaccgtgcac catgagcacg aatcctaaac gcaggttctc gagggcgcc ggttctttt gtcaagaccg gaggttctttt gtcaagaccg gaggttctttt gtcaagaccg gaggttctttt gtcaagaccg gaggcatcgc tggaggcac gctgatacc gcgtgcaacc gctgatacc gctgccagaa ggcgctaccg gagctgcacc gagcgcaccg gaactgtccg gagcaccg tactcggat gaagccggtc ggttcttttt gtcacttgcc gaaccgtcacc gctgccagcc gaactgtccg gagcacccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat ggcgatcccat gctgtgccac ccgtgatatt gctgaagagc gcgttttaccg tactcgccgc ccgattcgc tactcgccgc cccgattcgc tactcgccat cccgattcgc tactcgcat cccgattcgc tactcgcat cccgattcgc tactcgcat cccgattcgc tactcgcat cccgattcgcat cccgattcgcat cccgattcgcat cccgattcgcat cccgattcgc	120 180 240 300 360 420 480 540 660 720 780 840 990 1080 1140 1200 1260

taggggtett teecet agtteetetg gaaget gaaceecea cetgge tgcaaaggeg gcacaa atggetetee teaage tatgggatet gatetg aaacgtetag geecee c atg gae egg gag Met Asp Arg Glu	ctett gaagacaa cgaca ggtgeete accce agtgeeac cgtat teaacaag ggge eteggtge cegaa ecaegggg a atg gea gea	ac aacgtetgta etg cggccaaaag gt tgtgagttgg gg getgaaggat ac atgetttaca ac gtggttttec teg tge gga gg	gcgacccttt gcaccccacgtgtat aaga atagttgtgg aaag gcccagaagg tacctgtgtgtttagt cgactttgaaaaac acggc gcg gtt ttc	ggcagcg 1500 atacacc 1560 gagtcaa 1620 ccattg 1680 ggttaaa 1740 ataatac 1800 gta ggt 1849
ctg ata ctc ttg a Leu Ile Leu Leu T 20	_		-	
ctc ata tgg tgg t Leu Ile Trp Trp I 35	Leu Gln Tyr Ph			_
caa gtg tgg atc c Gln Val Trp Ile F 50				
atc ctc ctc acg t Ile Leu Leu Thr C . 65				
aaa atc ttg ctc g Lys Ile Leu Leu A		_		Gly
ata acc aaa gtg c Ile Thr Lys Val P 100				
tgc atg ctg gtg c Cys Met Leu Val A 115		a Gly Gly His		
ctc atg aag ttg g Leu Met Lys Leu A 130				
acc cca ctg cgg g Thr Pro Leu Arg A 145	Asp Trp Ala Hi		Arg Asp Leu Ala	
gca gtt gag ccc g Ala Val Glu Pro V 1	_			Thr
tgg ggg gca gac a Trp Gly Ala Asp T 180				
gtc tcc gcc cgc a Val Ser Ala Arg A 195		u Ile His Leu		

						cga Arg 215									2473
	_	_	_			ctt Leu		_			_				2521
	_			_	_	gag Glu			_		_			-	2569
				_		acc Thr	_	_		 	_			_	2617
						aag Lys									2665
						gtg Val 295									2713
				_		ttg Leu			_	_		_	-	-	2761
		-	_	_	_	cat His	-	_	-	_					2809
						cta Leu									2857
_			_			cca Pro			-	_					2905
				-	-	gtg Val 375	_		_	 -	_	_			2953
_		_		-		tct Ser	-	_		_			_	_	3001
						cct Pro									3049
						act Thr									3097

_	gcg Ala		_	_				_			_	_		_		3145
_	gcc Ala 450	_								_		_	_			3193
	gac Asp															3241
	atc Ile	_						_			_	_			_	3289
	gly aaa		_		_				_	_		_				3337
	tcg Ser															3385
	gct Ala 530														gga Gly	3433
	gtc Val															3481
	gga Gly	-							_							3529
	gjå aaa															3577
	ctc Leu	_		_	_							_	_	_		3625
	cgg Arg 610															3673
	gta Val															3721
tca Ser	gtg Val	atc Ile	gac Asp	tgc Cys 645	aat Asn	aca Thr	tgt Cys	gtc Val	acc Thr 650	cag Gln	aca Thr	gtc Val	gac Asp	ttc Phe 655	agc Ser	3769

											gtg Val					3817
											agg Arg					3865
								-			tcg Ser 700		-		-	3913
	_	-	_	_		-		_			tgt Cys	-				3961
											gct Ala				aca Thr	4009
											ttc Phe					4057
							_	_			ttg Leu		_		aag Lys :	4105
_	_		-					_	_	_	tac Tyr 780	_	_	_		4153
											gac Asp					4201
-					_		_	_			cca Pro	_			_	4249
		_		_	_				-		acc Thr					4297
											ctg Leu					4345
											gct Ala 860					4393
											agg Arg					4441

	agg Arg															4489
gat Asp	gag Glu	atg Met	gaa Glu 900	gag Glu	tgt Cys	gcc Ala	tca Ser	cac His 905	ctc Leu	cct Pro	tac Tyr	atc Ile	gaa Glu 910	cag Gln	gga Gly	4537
_	cag Gln		_	_				_	_	_			_	_		4585
	gcc Ala 930															4633
	cgg Arg															4681
	Gly aaa															4729
	ata Ile															4777
	acc Thr		His					Asn					Trp			4825
	caa Gln 1010	Leu					Āla					Val				4873
	gct Ala 5					Gly					Gly					4921
_	att Ile	-	Ala		$\mathtt{Tyr}$		_			Ala		_			Ala	4969
	aag Lys	_	_	Ser			_		Ser			_	_	Val		5017
	ctc Leu		Āla					Gly					Gly			5065
	gca Ala 1090	Ala					His					Glu				5113

			ct tog ogg ggt La Ser Arg Gly 1115		
		Pro Glu Ser As	ac gct gca gca sp Ala Ala Ala 130		Thr
			ag ctg ctg aag in Leu Leu Lys		
	Asn Glu Asp		ca tgc tcc ggc co Cys Ser Gly 116	Ser Trp	
	Trp Asp Trp		g ttg.act gat al Leu Thr Asp 1180		
			eg ccg gga gtc eu Pro Gly Val 1195		
		ys Gly Val Tr	gg cgg ggc gac p Arg Gly Asp 210		Met
			c acc gga cat le Thr Gly His		
	Arg Ile Val		cc tgt agt aac ir Cys Ser Asn 124	Thr Trp	
gga aca ttc Gly Thr Phe 1250	Pro Ile Asn	geg tac acc ac la Tyr Thr Th .255	eg ggc ccc tgc ir Gly Pro Cys 1260	acg ccc Thr Pro	tcc 5593 Ser
			gg cgg gtg gct pp Arg Val Ala 1275		
		al Gly Asp Ph	cc cac tac gtg ne His Tyr Val 190		Met
			ng gtt ccg gcc In Val Pro Ala		
ttc aca gaa Phe Thr Glu 1315	Val Asp Gly V	tg cgg ttg ca al Arg Leu Hi 1320	ac agg tac gct Ls Arg Tyr Ala 132!	Pro Ala	tgc 5785 Cys

		aca ttc ctg gtc ggg Thr Phe Leu Val Gly 1340	
- <b>-</b>	_	tgc gag ccc gaa ctg Cys Glu Pro Glu Leu 1355	
	Leu Thr Asp	ccc tcc cac att acg Pro Ser His Ile Thr 1370	
		tct ccc ccc tcc ttg Ser Pro Pro Ser Leu 1385	_
		tcc ttg aag gca aca Ser Leu Lys Ala Thr O : 140	Cys Thr Thr
_		ctc atc gag gcc aac Leu Ile Glu Ala Asn 1420	
		acc cgc gtg gag tca Thr Arg Val Glu Ser 1435	
	Ser Phe Glu	ccg ctc caa gcg gag Pro Leu Gln Ala Glu 1450	
		atc ctg cgg agg tcc Ile Leu Arg Arg Ser 1465	
		cgc ccg gat tac aac Arg Pro Asp Tyr Asn 148	Pro Pro Leu
		tac gtc cct cca gtg Tyr Val Pro Pro Val 1500	
		cct ccg ata cca cct Pro Pro Ile Pro Pro 1515	
	Leu Ser Glu	tct acc gtg tct tct Ser Thr Val Ser Ser 1530	
		age tee gaa teg teg Ser Ser Glu Ser Ser 1545	

agc ggc acg gca Ser Gly Thr Ala 1555	Thr Ala Ser F			
gcg gga tcc gac Ala Gly Ser Asp 1570				
gag ccg ggg gat Glu Pro Gly Asp 1585	_		Trp Ser Thr	-
gag gag gct agt Glu Glu Ala Ser			-	
aca ggc gcc ctg Thr Gly Ala Leu 1620	Ile Thr Pro C			Leu Pro
atc aat gca ctg Ile Asn Ala Leu 1635	Ser Asn Ser I			
gct aca aca tct Ala Thr Thr Ser 1650				
gac aga ctg cag Asp Arg Leu Gln 1665			Asp Val Leu	
atg aag gcg aag Met Lys Ala Lys				
gaa gcc tgt aag Glu Ala Cys Lys 1700	Leu Thr Pro P		_	Phe Gly
tat ggg gca aag Tyr Gly Ala Lys 1715	Asp Val Arg A			
atc cgc tcc gtg Ile Arg Ser Val 1730				
gac acc acc atc Asp Thr Thr Ile				
1745	Met Ala Lys A	175	_	1760

#### 68 / 93

gtt cgt		Glu Lys			Tyr					Thr		7177
cct cag	gcc gtg Ala Val 1795	atg ggo Met Gly	Ser Se	a tac er Tyr 100	gga Gly	ttc Phe	caa Gln	tac Tyr 1805	Ser	cct Pro	gga Gly	7225
cag cgg Gln Arg 1810	Val Glu							Lys				7273
atg ggc Met Gly 1825			Thr Ar				Ser					7321
aat gac Asn Asp						Gln					Ala	7369
ccc gaa		Gln Ala	_	_	Leu					Tyr		7417
Gly Gly aaa aac			Ser Ly						Tyr			7465
tgc cgc Cys Arg 1890	Ala Ser							Asn				7513
tgt tac Cys Tyr 1905			Ala Al				Ala					7561
tgc acg Cys Thr	_	-		-		Val	_		_	_	Ser	7609
gcg ggg		Glu Asp			Leu					Glu		7657
atg act Met Thr			Pro Pr						Pro			7705
gac ttg Asp Leu 1970	Glu Leu							Ser				7753
gat gca Asp Ala 1985			Val Ty				Arg					7801

;

#### 69 / 93

				gct Ala 2009	Ala				_	Arg					Asn	7849
				aac Asn )					Ala					Ala		7897
_		_	Met	act Thr				Ser				_	Gln		_	7945
		Lys	_	cta Leu	_	_	Gln				_	Cys				7993
	Pro			cta Leu		Gln					Leu					8041
				cat His 208	Ser					Glu					Ala	8089
	_			aaa Lys O			_	_	Pro	_	_	_		Arg		8137
			Ser	gtc Val				Leu					Gly		gct Ala	8185
_		Cys		aag Lys			Phe			-	-	Arg				8233
	Leu			atc Ile	_	Ala			_	_	Asp			_		8281
				tac Tyr 216	Ser					Tyr					Arg	8329
				tgg Trp O					Leu					Val		8377
			Tyr	cta Leu				Arg		acg	gga	gct a	aaaca	actco	ca	8427
ttt cca	tttt tctt	ttt agc	tttt: ccta	eteci	tt ti	tttt! ctag	ttec:	cti g aa	tttti	ttcc	ttt	tctt	tcc 1	tttgg	etttt gtggct cagaga	8547

<210> 8

<2	211> 212> 213>	DNA						
	100> ccago						•	6
<2 <2	210> 211> 212> 213>	63 DNA						
			tggcgcgccc	agatgttaac	cagatccatg	gcacactcta	gagtactgtc	60 63
<2 <2	210> 211> 212> 213>	33 DNA						
	100> Jgaat		aacagaccac	aacggtttcc	ctc			33
<2 <2	210> 211> 212> 213>	30 DNA						
	l00> gcgta		tggtattatc	gtgtttttca				30
<2 <2	210> 211> 212> 213>	45 DNA						
	100> catat		tctaatacga	ctcactatag	gccagccccc	gattg		45
<2 <2	210> 211> 212> 213>	45 DNA						
	cgcg		ttggtttttc	tttgaggttt	aggattcgtg	ctcat		45
<2 <2	110> 111> 112> 113>	36 DNA						
	.00> .aggg		tgattgaaca	agatggattg	cacgca			36
	10>							

<212> DNA <213> HCV	
<400> 15 gcatatgtta actcagaaga actcgtcaag aaggcgata	9
<210> 16 <211> 45 <212> DNA <213> HCV	
<400> 16 gcatatgaat totaatacga otcactatag gccagccccc gattg 4	5
<210> 17 <211> 30 <212> DNA <213> HCV	
<400> 17 acgcagaaag cgtctagcca tggcgttagt 3	0
<210> 18	
<400> 18 tcccggggca ctcgcaagca ccctatcagg 3	0
<210> 19 <211> 26 <212> DNA <213> HCV	
<220> <223> Label with FAM: fluorescence reporter dye	
<223> Label with TAMRA: Quencher dye	
<400> 19 tggtctgcgg aacgggtgag tacacc 2	6
<210> 20 <211> 45 <212> DNA <213> HCV	
<400> 20 gtggacgaat totaatacga otcactataa ocagococog attgg 4	5
<210> 21 <211> 27 <212> DNA <213> HCV	
<400> 21 ggaacgcccg tcgtggccag ccacgat 2	7

```
<210> 22
<211> 23
<212> DNA
<213> HCV
<400> 22
gtcgtcttct ctgacatgga gac
                                                                   23
<210> 23
<211> 27
<212> DNA
<213> HCV
<400> 23
gagttgctca gtggattgat gggcagc
                                                                   27
<210> 24
<211> 8638
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 24
accagecece gattggggge gacaetecae catagateae teceetgtga ggaactaetg 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gatcaacccg etcaatgeet ggagatttgg gegtgeecce 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
eggeegettg ggtggagagg etattegget atgaetggge acaacagaca ateggetget 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgaegggegt teettgegea getgtgeteg aegttgteae tgaageggga agggaetgge 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtateeat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacca ccaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggeteaa ggegegeatg ceegaeggeg aggatetegt egtgaeeeat ggegatgeet 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccqattcgc 1140
agcgcatcgc cttctatcgc cttcttgacg agttcttctg agttcgcgcc cagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtett teceeteteg ceaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440
agtteetetg gaagettett gaagacaaac aacgtetgta gegaecettt geaggeageg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatet gatetgggge eteggtgeae atgetttaea tgtgtttagt egaggttaaa 1740
aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
```

Me 1		EP A	g G	lu Me		la Al	la Se	er Cy		ly GI LO	Ly A	La Va	al Pl		al Gly L5	
										aag Lys						1897
										agg Arg						1945
										gjå aaa						1993
										cta Leu 75						2041
										atg Met						2089
										cac His						2137
_	_	_			_	_	_			cat His		-		_	_	2185
										tac Tyr						2233
		_		_						cta Leu 155		_				2281
										gag Glu						2329
			_				_		_	atc Ile	_	_		_		2377
										ctg Leu						2425
										cct Pro						2473
										atc Ile 235						2521

			gtc Val								2569
			gcg Ala								2617
			tca Ser								2665
			aat Asn								2713
		 _	tcc Ser 310	_		_		-	 _	_	 2761
			agg Arg								2809
			agc Ser								2857
			ggt Gly								2905
			gcc Ala								2953
			gag Glu 390								3001
	_	-	tcc Ser		_	 -	_	_	_		 3049
			cct Pro								3097
			caa Gln								3145
			ggt Gly								3193

										acc Thr			3241
		_				_			_	gac Asp	 	_	3289
										tgc Cys			3337
_	_			-				-	_	gac Asp 525			3385
										acg Thr			3433
										gct Ala			3481
		_					_			atc Ile			3529
										aag Lys			3577
										gct Ala 605			3625
										gga Gly			3673
										ggc Gly			3721
										gtc Val			3769
-	_	_				_	_			cca Pro			3817
										ggc Gly 685			3865

						cca Pro 695										3913
						tgc Cys										3961
	_		_			tca Ser	_		_		_					4009
						cag Gln										4057
						ata Ile										4105
_	_		_			ccc Pro 775		_	_	_		_	_	_		4153
						cca Pro										4201
						cct Pro										4249
						caa Gln										4297
						tgc Cys										4345
						ggc Gly 855										4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
						ccc Pro										4489
						gcc Ala										4537

Met Gln Leu 915	gcc gaa Ala Glu	Gln Phe						4585
aca gcc acc Thr Ala Thr 930								4633
tgg cgg acc Trp Arg Thr 945	_	_		_				4681
agc gġg ata Ser Gly Ile						Gly A		4729
gcg ata gca Ala Ile Ala	_			_		_	_	4777
acc acc caa Thr Thr Gln 99	His Thr	Leu Leu				Trp V		4825
gcc caa ctt Ala Gln Leu 1010			Ala Ala					4873
atc gct gga Ile Ala Gly 1025					Gly Lys			4921
gat att ttg Asp Ile Leu	Ala Gly							4969
	1045		ALG GLY	1050	GIY AIA		/al Ala 1055	
ttt aag gtc Phe Lys Val	atg agc	ggc gag	atg ccc	1050 tcc acc Ser Thr	gag gac	ctg g	.055 gtt aac	5017
	atg agc Met Ser 1060 gct atc Ala Ile	ggc gag Gly Glu ctc tcc Leu Ser	atg ccc Met Pro 1065 cct ggc	tcc acc Ser Thr	gag gac Glu Asp gtc gtc	ctg g Leu V 1070 ggg g Gly V	1055 gtt aac Val Asn gtc gtg	5017 5065
Phe Lys Val	atg agc Met Ser 1060 gct atc Ala Ile 5	ggc gag Gly Glu ctc tcc Leu Ser	atg ccc Met Pro 1065 cct ggc Pro Gly 1080 cac gtg His Val	tcc acc Ser Thr gcc cta Ala Leu	gag gac Glu Asp gtc gtc Val Val 1085	ctg g Leu V 1070 ggg g Gly V	oft aac Val Asn Val Val Val Gtg	
Cta ctc cct Leu Leu Pro 1079 tgc gca gcg Cys Ala Ala	atg agc Met Ser 1060 gct atc Ala Ile 5 ata ctg Ile Leu aac cgg	ggc gag Gly Glu  ctc tcc Leu Ser  cgt cgg Arg Arg 1095	atg ccc Met Pro 1065 cct ggc Pro Gly 1080 cac gtg His Val	tcc acc Ser Thr gcc cta Ala Leu ggc cca Gly Pro	gag gac Glu Asp gtc gtc Val Val 1085 ggg gag Gly Glu 1100 cgg ggt Arg Gly	ctg g Leu V 1070 ggg g Gly V ggg g Gly A	gtt aac Val Asn Val Val Vat Yal Val Val Val Val Val	5065

cag atc Gln Ile		Ser Leu			Gln	_	-	_		Leu		5257
cag tgg Gln Trp				Thr					Ser			5305
aga gat Arg Asp 1170	Val Trp							Asp				5353
tgg ctc Trp Leu 1185	_	-	Leu Pro	_	Leu	_	Gly	_				5401
tca tgt Ser Cys						Arg					Met	5449
caa acc Gln Thr		Pro Cys			Ile					Ьуs		5497
ggt tcc Gly Ser				Arg					Thr			5545
gga aca Gly Thr 1250	Phe Pro							Суз				5593
ccg gcg Pro Ala 1265			Arg Ala		Trp		Val					5641
tac gtg Tyr Val						His					Met	5689
acc act Thr Thr	_	Val Lys		_	Gln	_	_	_		Glu		5737
ttc aca Phe Thr				Leu					Pro			5785
aaa ccc Lys Pro 1330	Leu Leu							Gly				5833
tac ctg Tyr Leu 1345			Leu Pro		Glu		Glu					5881

gtg ctc act tcc Val Leu Thr Ser			His Ile Thr		5929
gct aag cgt agg Ala Lys Arg Arg 138	g Leu Ala Arg				5977
tca gct agc cag Ser Ala Ser Glr 1395				Cys Thr Thr	6025
cgt cat gac too Arg His Asp Ser 1410		Asp Leu Ile			6073
cgg cag gag ato Arg Gln Glu Met 1425		_		-	6121
gta gta att ttg Val Val Ile Lev			Gln Ala Glu		6169
agg gaa gta too Arg Glu Val Sen 146	. Val Pro Ala				6217 
cct cga gcg atg Pro Arg Ala Met 1475				Pro Pro Leu	6265
tta gag tcc tgg Leu Glu Ser Trp 1490		Asp Tyr Val			6313
tgt cca ttg ccg Cys Pro Leu Pro 1505					6361
aag agg acg gtt Lys Arg Thr Val	-	_	Val Ser Ser		6409
gag ctc gcc aca Glu Leu Ala Thi 154	Lys Thr Phe				6457
agc ggc acg gca Ser Gly Thr Ala 1555	a acg gcc tct a Thr Ala Ser	cct gac cag Pro Asp Gln 1560	ccc tcc gac Pro Ser Asp 1565	Asp Gly Asp	6505
gcg gga tcc gad Ala Gly Ser Asp 1570		Tyr Ser Ser			6553

gag ccg ggg gat Glu Pro Gly Asp 1585	•		Trp Ser Thr	
gag gag gct agt Glu Glu Ala Ser			_	- <del>-</del>
aca ggc gcc ctg Thr Gly Ala Leu 1620	Ile Thr Pro			Leu Pro
atc aat gca ctg Ile Asn Ala Leu 1635	Ser Asn Ser		_	-
gct aca aca tct Ala Thr Thr Ser 1650		Ser Leu Arg Glr		
gac aga ctg cag Asp Arg Leu Gln 1665			Asp Val Leu	
atgʻaag gcg aag Met Lys Ala Lys				
gaa gcc tgt aag Glu Ala Cys Lys ' 1700	Leu Thr Pro		-	Phe Gly
tat ggg gca aag Tyr Gly Ala Lys 1715	Asp Val Arg			
atc cgc tcc gtg Ile Arg Ser Val 1730		Leu Leu Glu Asp		
gac acc acc atc Asp Thr Thr Ile 1745			Cys Val Gln	
aag ggg ggc cgc Lys Gly Gly Arg				
gtt cgt gtg tgc Val Arg Val Cys 1780	Glu Lys Met			Thr Leu
cct cag gcc gtg Pro Gln Ala Val 1795	Met Gly Ser			

	gag ttc ctg Glu Phe Leu			a Lys Lys	
	gca tat gad Ala Tyr Asp 183	Thr Arg Cy			
	cgt gtt gag Arg Val Glu 1845				
	aga cag gcc Arg Gln Ala 1860		r Leu Thr Gl		Tyr Ile
	ctg act aat Leu Thr Asn 5				
	agc ggt gta Ser Gly Val			y Asn Thr	
-	aag gcc gct Lys Ala Ala 191	Ala Ala Cy			
	ctc gta tgc Leu Val Cys 1925				
	caa gag gac Gln Glu Asp 1940		r Leu Arg Ala		Glu Ala
	tac tct gcc Tyr Ser Ala 5				
	ttg ata aca Leu Ile Thr	_		l Ser Val	
	ggc aaa agg Gly Lys Arg 199	Val Tyr Tyr			
	cgg gct gcg Arg Ala Ala 2005				
	ggc aac atc Gly Asn Ile 2020		r Ala Pro Th		Ala Arg

			ctt cta gct cag Leu Leu Ala Gln 2045	•
		Gln Ile Tyr	ggg gcc tgt tac Gly Ala Cys Tyr 2060	
			cga ctc cac ggc Arg Leu His Gly 2075	
			gag atc aat agg Glu Ile Asn Arg O	
	Lys Leu Gly		ttg cga gtc tgg Leu Arg Val Trp 211	Arg His
			tcc cag ggg ggg Ser Gln Gly Gly 2125	
		Phe Asn Trp	gca gta agg acc Ala Val Arg Thr 2140	
			ttg gat tta tcc Leu Asp Leu Ser 2155	
			tat cac agc ctg Tyr His Ser Leu O	_
	Trp Phe Met		ctc cta ctt tct Leu Leu Leu Ser 219	Val Gly
gta ggc atc tat Val Gly Ile Tyr 2195			acggggaget aaac	actcca 8427
ttttttttt tttt	steett ttttt gteacg getag	tteet etttt ctgtg aaaggt	tttc tttttttttt ttcc ttttctttcc ccgt gagccgcttg	tttggtggct 8547
<210> 25 <211> 8638 <212> DNA <213> HCV			·	
<220> <221> CDS <222> (1802)	(8407)			

```
<400> 25
accageceee gattggggge gacacteeae catagateae teccetgtga ggaactactg 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gatcaacccg etcaatgeet ggagatttgg gegtgeecce 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgcc ggttcttttt gtcaagaccg 540
acctgteegg tgeectgaat gaactgeagg acgaggeage geggetateg tggetggeea 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getacetgee 780
cattegacca ccaagegaaa categcateg agegageacg tacteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggetcaa ggegegeatg eccgaeggeg aggatetegt egtgaeceat ggegatgeet 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggeggega atgggetgae egetteeteg tgetttaegg tategeeget eeegattege 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee cagatgttaa 1200
cagaccacaa eggttteeet etagegggat caatteegee ecceeecta aegttaetgg, 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc 1440
agttectetq qaagettett gaagacaaac aacgtetgta gegaceettt geaggeageg 1500
gaacceccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggetetee teaagegtat teaacaaggg getgaaggat geceagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
                                       10
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg
                                                                  1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
             20
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                  1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
         35
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                  1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
                         55
atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc
                                                                  2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
                                                                  2089
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca
                                                                  2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
            100
                                105
```

			cgg Arg										2185
			gcc Ala										2233
			gac Asp										2281
			gtc Val 165										2329
 	_	_	acc Thr			_		_		_			2377
			agg Arg										2425
_		-	gjå aaa		_					_	_		2473
_	_	_	ggc				_			_			2521
			cag Gln 245										2569
			ctg Leu			_	_		 	_		-	2617
			Gly	-	_			_		_			2665
	_		acc Thr			-	_	_	_				2713
			cgt Arg										2761
			acg Thr 325										2809

										ccc Pro					2857
										tcg Ser					2905
										gtt Val 380					2953
										atg Met					3001
										cag Gln					3049
										agc Ser					3097
_			_	_			_			gtc Val	_		_		3145
										tct Ser 460					3193
	_				_		_			atc Ile		_			3241
		_					_	-		gcc Ala	_			_	3289
			_		_			_	_	gag Glu	_				3337
										ctg Leu					3385
										gct Ala 540					3433
										gtg Val					3481

act c		_						_							3529
aag g Lys G															3577
gag d Glu I															3625
tac o				_	_	_				-		_	_		3673
gtc g Val V 625															3721
tca s Ser V															3769
ctg c															3817
gtg t															3865
att t Ile 1							_			_		_		_	3913
tcc t Ser S 705															3961
ctc a Leu 1	_		_			_		_							4009
cca c Pro C		_		gtc	_	 _		ctg					ggc	_	4057
ttt a Phe 1						_	_			_		_		_	4105
cag g Gln A			_				_	_	_		_	_	_		4153
tgc g Cys I 785	_		_	_	_			_		_		_		_	4201

												acg Thr				4249
												aca Thr				4297
												gag Glu 845				4345
												ctg Leu				4393
												atc Ile				4441
												tac Tyr				4489
_		_	_		_	_						atc Ile	_	_		4537
_	_		-	-				_	_	_		ggg Gly 925	_	_		4585
												gtg Val				4633
				_	_				_		_	tgg Trp				4681
												cct Pro				4729
												acc Thr				4777
acc Thr	acc Thr	caa Gln 995	His	acc Thr	ctc Leu	ctg Leu	ttt Phe 1000	Asn	atc Ile	ctg Leu	gl <sup>à</sup> aaa	gga Gly 1005	${\tt Trp}$	gtg Val	gcc Ala	4825
		Leu					Ala					gta Val )				4873

	Ala					ggc Gly					Gly					4921
					Tyr	gga Gly				Ala					Ala	4969
				Ser		gag Glu			Ser					Val		5017
			Āla			tcc Ser		Gly					Gly			5065
		Ala				cgg Arg 1099	His					Glu				5113
	Trp					ata Ile )					Arg					5161
					Val	cct Pro				Āla					Thr	5209
_				Ser		acc Thr			Gln	_	~	_	~~	Leu		5257
_			Asn		_	tgc Cys		Thr		_			Ser			5305
_	_	Val		_		ata Ile 1175	Cys	_		_		Āsp		_	_	5353
	Leu					ctg Leu )					Gly					5401
					Tyr	aag Lys				Arg					Met	5449
				Pro	-	gga Gly	_	_	Ile					Lys		5497
			Arg			gjå aaa		Arg					Thr			5545

Gly T		Phe					Tyr					Cys	acg Thr			5593
ccg g Pro A 1265						Arg					Val					5641
tac g Tyr V					Arg					His					Met	5689
acc a Thr T		-		Val	_	_	_	_	Gln	_	_	_		Glu		5737
ttc a Phe T	Chr (		Val					Leu					Pro			5785
aaa c Lys P 1		Leu					Val					Gly				5833
tac c Tyr P 1345	_	_			_	Leu		_			Glu	_	_	_	_	5881
gtg c					Leu					His					Thr .	5929
	Leu '	Thr cgt	Ser agg	Met 1365 ctg Leu	Leu 5 gcc	Thr	Asp	Pro tct	Ser 1370 ccc Pro	His CCC	Ile	Thr	Ala	Glu 137! agc Ser	Thr .	5929 5977
Val L	Leu :	Thr cgt Arg	agg Arg 1380 cag	Met 1365 ctg Leu ) ctg	Leu gcc Ala tct	Thr agg Arg	Asp gga Gly cct	tct Ser 1385 tcc Ser	Ser 1370 ccc Pro ttg	His ccc Pro	Ile tcc Ser	Thr ttg Leu aca	gcc Ala 1390 tgc Cys	Glu 137! agc Ser	tca Ser	
yal L gct a Ala L tca g Ser A cgt c	Leu :	Thr cgt Arg agc Ser 1395 gac Asp	agg Arg 1380 cag Gln	Met 1365 ctg Leu ctg Leu	gcc Ala tct Ser	Thr agg Arg gcg Ala gct	gga Gly cct Pro 1400 gac	tct Ser 1385 tcc Ser	Ser 1370 ccc Pro ttg Leu	His ccc Pro aag Lys	tcc Ser gca Ala	ttg Leu aca Thr 1409 aac	gcc Ala 1390 tgc Cys	agc ser act Thr	tca Ser acc Thr	5977
yal L gct a Ala L tca g Ser A cgt c	aag-obys det de la	Thr cgt Arg agc Ser 1395 gac Asp	agg Arg 1380 cag Gln tcc ser	Met 1365 ctg Leu ctg Leu ccg Pro	gcc Ala tct ser gac Asp	agg Arg gcg Ala gct Ala 1415 aac	gga Gly cct Pro 1400 gac Asp	tct Ser 1385 tcc Ser ctc Leu	Ser 1370 CCC Pro ttg Leu atc Ile	ccc Pro aag Lys gag Glu	tcc Ser gca Ala gcc Ala 1420 gag Glu	ttg Leu aca Thr 1409 aac Asn	gcc Ala 1390 tgc Cys ctc Leu	agc ser act Thr ctg Leu	tca Ser acc Thr tgg Trp	5977 6025
yal L gct a Ala L tca g Ser A cgt c Arg H 1 cgg c Arg G	aag-obys de	Thr  cgt Arg  agc ser 1395 gac Asp gag Glu att	agg Arg 1380 cag Gln tcc ser atg Met	ctg Leu ctg Leu ctg Leu ccg Pro	gcc Ala  tct Ser  gac Asp  ggg Gly 1430  tct Ser	Thr agg Arg gcg Ala gct Ala 1415 aac Asn	gga Gly cct Pro 1400 gac Asp atc Ile	tct ser 1385 tcc ser ctc Leu acc Thr	ser 1370 ccc Pro ttg Leu atc Ile cgc Arg	CCC Pro  aag Lys  gag Glu  gtg Val 1435  caa Gln	tcc ser gca Ala gcc Ala 1420 gag Glu	ttg Leu aca Thr 1409 aac Asn tca Ser	gcc Ala 1390 tgc Cys ctc Leu gag Glu	agc ser act Thr ctg Leu aat Asn	tca ser acc Thr tgg Trp aag Lys 1440 gag Glu	5977 6025 6073

cct Pro	cga Arg	gcg Ala 1475	Met	ccc Pro	ata Ile	tgg Trp	gca Ala 1480	Arg	ccg Pro	gat Asp	tac Tyr	aac Asn 1489	cct Pro	cca Pro	ctg Leu	6265
		Ser					Asp					Val	gta Val			6313
	Pro					Lys					Pro		cca Pro			6361
					Leu					Val			gcc Ala		Ala	6409
				ГÀз					Ser				gcc Ala 1550	Val		6457
_		_	Ala	_	_			Asp	_			_	gac Asp		_	6505
		Ser					Tyr					Pro	ctt Leu			6553
	Pro					Leu					Trp		acc Thr			6601
					Asp					Ser			tac Tyr		Trp	6649
				Ile					Ala				aag Lys 1630	Leu		6697
		-	Leu	_			_	Leu	-				ttg Leu 5	-		6745
Ala		Thr		Arg	Ser	_	Ser	_		Gln	_	Lys	gtc Val			6793
	Arg					Asp					qaA		ctc Leu			6841
					Ser					Lys			tcc Ser		Glu	6889

			ct aaa ttt ggc 69: er Lys Phe Gly 1710	37
Lys Asp Val		tcc agc aag g Ser Ser Lys A 1		85
		gaa gac act g Glu Asp Thr G 1740		33
	Lys Asn Glu		te caa eca gag 709 al Gln Pro Glu 1760	81
		atc gta ttc c Ile Val Phe P 1770	ca gat ttg ggg 71: ro Asp Leu Gly 1775	29
			tc tcc acc ctc 71° al Ser Thr Leu 1790	77
 Val Met Gly		Gly Phe Gln T	ac tct cct gga 72 yr Ser Pro Gly 805	25
		tgg aaa gcg a Trp Lys Ala L 1820		73
	Thr Arg Cys		cg gtc act gag 73: hr.Val Thr Glu 1840	21
		tac caa tgt t Tyr Gln Cys C 1850		69
		ctc aca gag c Leu Thr Glu A 5		17
 Leu Thr Asn		Gln Asn Cys G	gc tat cgc cgg 740 ly Tyr Arg Arg 885	65
			at acc ctc aca 75 sn Thr Leu Thr	13
	Ala Ala Cys		ag ctc cag gac 750 ys Leu Gln Asp 1920	61

tgc Cys	acg Thr	atg Met	ctc Leu	gta Val 192	Суз	gga Gly	gac Asp	gac Asp	ctt Leu 1930	Val	gtt Val	atc Ile	tgt Cya	gaa Glu 1935	Ser	7609
gcg Ala	ej aaa	acc Thr	caa Gln 1940	Glu	gac Asp	gag Glu	gcg Ala	agc Ser 194!	Leu	cgg Arg	gcc Ala	ttc Phe	acg Thr 1950	Glu	gct Ala	7657
			Tyr			ccc Pro		Gly					Pro			7705
		Glu				tca Ser 1975	Сув					Ser				7753
	Ala					gtg Val )					Arg					7801
					Ala	tgg Trp				Arg					Asn	7849
				Asn		atc Ile			Ala					Ala		7897
_		_	Met			ttc Phe		Ser				_	Gln	-		7945
	_	Lys	_		-	tgt Cys 2055	Gln				_	Cys				7993
	Pro		_			cag Gln )				_	Leu				_	8041
					Ser	tac Tyr				Glu					Ala	8089
				Lys		gjå aaa			Pro					Arg		8137
			Ser			gct Ala		Leu					Gly			8185
		Cys				ctc Leu 2135	Phe					Arg				8233

	Leu					Ala					Asp				tgg Trp 2160	8281
ttc Phe	gtt Val	gct Ala	ggt Gly	tac Tyr 2165	Ser	ely aaa	gga Gly	gac Asp	ata Ile 2170	Tyr	cac His	agc Ser	ctg Leu	tct Ser 217		8329
				Trp		atg Met			Leu					Val	gjà aaa	8377
			$\mathtt{Tyr}$			ccc Pro		Arg		acg	gggag	gct a	aaac	actć	ca	8427
cca	tttt tctta	agc o	ctag	ctect gtca	t to	tttt	ttee! ctgts	t ctt g aaa	tttt	ttcc	ttt	tctti	tcc	tttg	ttttt gtggct cagaga	8547